University of Alberta

The Problem of Exportation for Dynamically Holistic Systems

by

Li Li ©

A thesis submitted to the Faculty of Graduate Studies and Research
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

Department of Philosophy

Edmonton, Alberta
Spring 2006

Library and
Archives Canada

Published Heritage
Branch

395 Wellington Street
Ottawa ON K1A 0N4
Canada

Bibliothèque et
Archives Canada

Direction du
Patrimoine de l'édition

395, rue Wellington
Ottawa ON K1A 0N4
Canada

Your file Votre référence
ISBN: 0-494-14008-9
Our file Notre référence
ISBN: 0-494-14008-9

NOTICE:
The author has granted a non-exclusive license allowing Library and Archives Canada to reproduce, publish, archive, preserve, conserve, communicate to the public by telecommunication or on the Internet, loan, distribute and sell theses worldwide, for commercial or non-commercial purposes, in microform, paper, electronic and/or any other formats.

The author retains copyright ownership and moral rights in this thesis. Neither the thesis nor substantial extracts from it may be printed or otherwise reproduced without the author's permission.

AVIS:
L'auteur a accordé une licence non exclusive permettant à la Bibliothèque et Archives Canada de reproduire, publier, archiver, sauvegarder, conserver, transmettre au public par télécommunication ou par l'Internet, prêter, distribuer et vendre des thèses partout dans le monde, à des fins commerciales ou autres, sur support microforme, papier, électronique et/ou autres formats.

L'auteur conserve la propriété du droit d'auteur et des droits moraux qui protège cette thèse. Ni la thèse ni des extraits substantiels de celle-ci ne doivent être imprimés ou autrement reproduits sans son autorisation.

In compliance with the Canadian Privacy Act some supporting forms may have been removed from this thesis.

While these forms may be included in the document page count, their removal does not represent any loss of content from the thesis.

Conformément à la loi canadienne sur la protection de la vie privée, quelques formulaires secondaires ont été enlevés de cette thèse.

Bien que ces formulaires aient inclus dans la pagination, il n'y aura aucun contenu manquant.

Canada

for Grant Weigao Wang & my family

Abstract

In this thesis, I address a problem that is long-standing in the philosophy of science, namely, the problem of "exportation"—how descriptions of idealized models can be used to describe real systems. I argue that current accounts of exportation are based on assumptions of some kinds of "atomism" that do not apply even in classical mechanics. Classical mechanics claims that the Hamiltonian and the motion of a nonseparable system cannot be decomposed into the Hamiltonians and motions of its subsystems. This feature of nonseparable systems I call dynamical holism. A consequence of dynamical holism is that the descriptions of dynamically holistic systems cannot be approximated by the descriptions of idealized systems, even when the disturbances approach zero. This raises problems for the current accounts of exportation as they require the approximation relation between the descriptions of idealized and real systems to hold when the disturbances approach zero.

I argue that we can solve the problem of exportation for dynamically holistic systems by employing the methods of singular perturbation theory. These methods provide us with ways to idealize and divide dynamically holistic systems, and thereby make exportation of idealized models to dynamically holistic systems possible. Singular perturbation theory shows that we need to introduce multiple scales into the descriptions of dynamically holistic systems and idealize them differently from the traditional idealization, that is, instead of idealizing them to systems without disturbances, we need to idealize them to systems with averaged perturbations, so that the idealized models can

be exportable as approximations to the dynamically holistic systems. It also shows that the traditional idealized models can be exported to describe the motion of a dynamically holistic system at one scale. This new account of exportation establishes that we do not have to restrict our exportations to "atomistic" cases as the traditional accounts do, but we do need to use modified idealized models, i.e., idealized models at different scales, to describe the behaviors of dynamically holistic systems.

Acknowledgements

I would like to thank my supervisors Alex Rueger and David Sharp for their guidance and for helping me refine my arguments. They have been patient and given so much extraordinary help to me. I also want to thank my colleague Patrick McGivern for his precious friendship and help. He not only discussed with me the problems I dealt with, helped me organize the thesis and clarify my ideas and expressions, he also has been always encouraging and supportive. I am also very grateful to Dr. Michael Li for helping me get a more precise understanding of the mathematical concepts. I also thank him and other members of the examining committee (besides my supervisors) – Dr. Margaret Morrison and Dr. Bruce Hunter for reading and liking the thesis. Special thanks go to Dr. Morrison for her very encouraging comments about the thesis. I am also grateful to Anita, Wendy and Dr. Burch for their friendly and kind help with my administrative needs during my study in the Department of Philosophy.

Many friends – especially Rev. Phil Taylor, Lorraine, Titus, Mary, Maureen, Harold, Hui, Jinglin and Sonya – have given me lots of encouragement and support and I am very grateful for that. Many thanks to my parents Shuhuai and Shuying for their love and support.

Most of all, I am very grateful to my husband Grant Wang for all his support, encouragement and help with the thesis and at home. If you did not insist that I had to finish the thesis, I might never have got it done. I am also grateful to my daughter Lena for her understanding and help at home and to my little son Christopher for making me relaxed while I was stressed.

Table of contents

Introduction 1

Chapter One. Dynamical holism in classical mechanics 8

Introduction 8

1.1. Basics of Hamiltonian Mechanics 11

1.2. The nonseparability of Hamiltonians and the nonintegrability of Hamiltonian systems 20

1.3. The notion of dynamical holism defined in terms of nonseparability 27

1.4. Some possible objections to the notion of dynamical holism 29

(a) Objection: dynamical holism is merely an epistemic problem 29

(b) Objection: dynamical properties supervene on states 40

(c) Objection: dynamical holism presumes the reality of "imaginary" subsystems 43

Conclusion 48

Chapter Two. Nickles nonreducibility of dynamically holistic systems 50

Introduction 50

2.1. The notion of Nickles reduction 52

2.2. Some basics concepts of perturbation theory 54

2.3. The distinction between regular and singular perturbation problems and the Nickles nonreducibility of singular perturbation problems 64

2.4. Nonseparable systems, can be treated only by singular perturbation theory 72

2.5. Qualitative irreducibility of Nickles nonreducible systems 75

Conclusion 80

Chapter Three. The conflict between dynamical holism and the current views of scientific methods 81

Introduction 81

3.1. McMullin's approximate truth view of scientific methods 83

3.2. Cartwright's account of scientific methods 88

(a) Cartwright's tendency view 90

(b) The conflict between Cartwright's tendency view and dynamical holism 93

(c) The measurability problem of Cartwright's view 101

3.3. Rueger and Sharp's structural stability view of scientific methods 106

Conclusion 109

Chapter Four. A solution to the exportation problem in dynamical holism 111

Introduction 111

4.1. Singular perturbation methods 114

(a) An example of a singular perturbation problem with an integrable Hamiltonian 115

(b) Solving the equations of dynamically holistic systems by singular perturbation methods 125

4.2. The conditions for the asymptotic solutions of singular perturbation methods to be valid 134

4.3. An account of idealization and exportation in a dynamically holistic world 137

(a) Multiple scales are necessary for obtaining approximate solutions for the dynamically holistic systems 138

(b) Exportations and idealizations in dynamically holistic cases 140

(c) A new account of traditional idealized models 154

Conclusion 157

Appendix 160

Bibliography 162

Introduction

Theories and models in science are usually taken to be idealizations. The real world rarely displays simple order or regular patterns of behavior. Although we can produce such order and behavior in the laboratory, it is clear that laboratory conditions are very special circumstances: scientists prepare systems by simplifying them, eliminating or ignoring 'non-relevant' factors and shielding off 'disturbing' factors. It is under these special conditions that we test our theories or models which themselves describe idealized systems. Without such idealization techniques, a scientific description of the world seems hardly possible. Modern science, as inaugurated by Galileo, is based on these techniques. Presumably, if Galileo had not made, or at least thought in terms of, all kinds of idealizations in his free-fall experiments, he would not have been able to describe the motion of bodies by the law of free fall. Similarly, Newton had to assume that the universe contains only two bodies to derive Kepler's laws. This is why scientific laws always contain provisos such as "if the system is isolated…" or "if the gas is ideal…".

The use of idealizations in scientific theories gives rise to a problem that has received a great deal of attention. Cartwright (1989) vividly describes it as the problem of "exportation": how can idealized theories or models be used to give us knowledge about the real world, as opposed to the specially prepared world in the laboratory? If the idealized theories have to 'falsify' the real world, how can they be used to arrive at correct descriptions of it?

The problem has been approached in a variety of ways. McMullin (1985), for instance, tries to connect idealizations with real systems in terms of approximations. He thinks that idealized laws can be taken as approximate descriptions of the behavior of real systems in the world. Cartwright herself (1989), by contrast, argues that the exportation problem cannot be solved by invoking approximations; in her view, we have to understand idealized models as describing dispositions or 'tendencies' that certain factors have. She argues that only these tendencies remain stable when we go from the idealized systems to real systems, not the manifested behavior of the systems. More recently Rueger and Sharp (1998) proposed that the exportation of idealized theories to real systems is justified by the structural stability of the theories: the idealized theories describe the structures or patterns of the behavior of both the idealized and the real systems. Although the real systems are perturbations of the idealized systems, the perturbations do not change the patterns of behavior described by the idealized theories.

Such approaches depend on more or less strong ontological assumptions about the world, as especially Cartwright has made clear. In her case, she concludes from both an analysis of the way we 'export' scientific claims from idealized systems to systems that are considerably different and the fact that we are often successful in this, that the world must be 'atomistic' in the sense that the factors involved in an idealized description tend to produce the same effect independently of other factors in non-ideal cases. Only when the factors involved in idealized descriptions are atomistic can the idealized theories be exported to real systems which are typically characterized by the presence of many different factors. If the world were not atomistic in this sense, the factors would not produce the same effect in the real systems as in the ideal system. Our practice of exporting idealized theories to real systems would then usually not result in successful predictions. Conversely, only in an atomistic world can we explain the success of our methods of idealization and exportation.

There are, however, indications, coming from science itself, that this assumption of atomism cannot be correct. Recently, some philosophers such as Howard (1989), Healey (1991), Maudlin (1998), and Teller (1986) have argued that quantum mechanics implies a form of 'holism' about the world which seems to be in conflict with the atomistic view of the world. A characterization of this form of holism had already been given by Einstein in his discussion of the EPR argument that Einstein and his co-writers use to show that quantum theory is inadequate: the world is holistic if it is such that it does not allow us to assign independent states to the physical objects situated in different parts of space. This means that the holistic state of a composite system would not be reducible to the independent states of its parts. Insofar as this sort of 'state holism' is a consequence of quantum mechanics, Einstein thought that this theory could not be a correct view of the world. If the world indeed were state holistic, he argued, it would be impossible to formulate and test the laws of physics; physical theorizing about the world would be impossible (M. Born, 1971, pp.170-171). So, for Einstein, if the world were not atomistic, we would in effect not be able to know anything about the world, not even the idealized laws.

Despite Einstein's opposition, of course, quantum theory has been "precisely formulated and rigorously tested" (ibid., p.49). This seems to indicate that we *can* know the laws even if the world is state holistic. The question then is, how is this possible? According to quantum theory, the whole universe is in an entangled state. This means that to know the state of any object in the world, strictly speaking, we would need to know the state of the whole universe. But as some philosophers argue, it is logically impossible to know the quantum state of the whole universe.[1] Moreover, as Scheibe (1995) argues, even if we could know the state of the universe, we still could not determine the states of the objects in the world, which are the subsystems of the universe, since the states of the subsystems are dependent on each other. This means that in order to know the state of any system in the universe which is in an entangled state, we would have to know the state of the universe and the states of all the other subsystems first. This indicates that in principle, it is impossible to know the state of any system in the quantum state holistic universe.

In this thesis, I will ignore such problems arising from quantum mechanics. But even if we focus on a 'classical' world, we still encounter problems associated with holism. Although *classical mechanics* does not have the problem of quantum state holism, some philosophers, such as Scheibe (1995), argue that it implies a different sort of holism, which I call 'dynamical holism'. In contrast to the holism concerning *states* of systems in quantum mechanics, dynamical holism is a feature of the *motion* or dynamical evolution of systems. This dynamical evolution is encoded in a system's Hamiltonian and the dynamical kind of holism can be characterized in terms of 'nonseparable' Hamiltonians. A system has a nonseparable Hamiltonian when its Hamiltonian cannot be decomposed into the Hamiltonians of its subsystems (even when the notion of a "subsystem" is broadly construed, in a sense to be made clear in Section 1.1). For example, the Hamiltonian of a system of two interacting particles sometimes is not the sum of the Hamiltonians of the particles as they are in isolation. In such cases, we can write it formally as: $H = H_1 + H_2 + H_{\text{int}} \neq H_1 + H_2$. If a composite system has a nonseparable Hamiltonian, the system's time evolution would not be the sum of and cannot be determined by the independent time evolutions of the subsystems (again, even broadly construed). Rather, the time evolutions of the interacting subsystems would be inextricably linked.

---

[1] See Breuer, 1995.

I will argue that this form of holism causes problems for the above mentioned accounts of idealization and exportation of scientific knowledge. If a system *were* dynamically holistic, and if we assume that we *do* have knowledge of relevant idealized models or laws, then these accounts of the success of exportation to such a system face principled difficulties. I will show (in Chapter Three) that if we use perturbation theory as the tool to describe exportation, the ways in which McMullin, Cartwright, and Rueger and Sharp try to justify exportation all rely (in different respects) on the assumption of atomism resulting from the application of *regular* perturbation methods. These methods, roughly, assume that the descriptions of real systems go over into the descriptions of the idealized systems in the limit of vanishing perturbations. Thus it is assumed that the effect of some factor appearing in the idealized system is separable from the effects of other perturbing factors. But this assumption cannot be upheld in the case of dynamically holistic systems.

My strategy in Chapter Four is to assume that a real system is correctly described by some equations of motion of a dynamically holistic kind (which we pretend not to know) and to explore under what conditions an idealized version of these equations can be used to approximate the behavior of the real system. (A further assumption is, of course, that the interactions within such a system or with its environment are weak enough so as to allow the application of perturbation methods to solve the equations of motion.) For dynamically holistic systems, it turns out that the regular approach leads to solutions that are not valid as approximations to the behavior of the real systems. Instead, *singular* perturbation methods have to be used. These methods enable us to find a kind of modification of the solutions for the idealized systems that are valid approximations of the solutions for the dynamically holistic systems. It is this use of singular perturbation theory that constitutes the essential difference in dealing with the question of idealization and exportation in a dynamically holistic world compared to an atomistic world. Several consequences follow from this difference.

In contrast to regular perturbation problems, we have in the dynamically holistic case a qualitative and/or quantitative difference between the solutions for the real systems and those of the idealized system -- even in the case of vanishing perturbations. The solutions of the dynamically holistic systems are not only quantitatively different from the solutions for the idealized systems, but also describe very different behavior patterns than those for the idealized systems. This is the basic reason why the above approaches to idealization and exportation are inadequate when applied to dynamically holistic systems and need to be replaced by a different view.

I think the account of exportation in dynamically holistic cases should be based on singular perturbation methods. Solving the equations of motion for systems that show the characteristics of dynamical holism (nonseparable Hamiltonians) by singular perturbation methods requires us to introduce different spatial and/or temporal scales into the approximate solutions of the equations. Thus, even though we start with a description of the real system at one such level or scale, in order to get (approximate) knowledge of the system's behavior, we are forced to divide reality up into a hierarchy of two or more levels (micro and macro, fast and slow scales). Only in this way can the behavior of the dynamically holistic system be approximately described. This is because the perturbations produce motions at other scales than the scale at which the motion of the idealized system is described. Without introducing the other scales, these motions cannot be correctly captured. I argue that a sort of 'dynamical atomism', at each scale or level of the solution, can be reestablished even in a (weakly) dynamically holistic world. It is this dynamical atomism at each scale that makes the exportation of idealized descriptions to dynamically holistic systems possible. The differences between this sort of atomism and the kind postulated by Cartwright are discussed at the end of Chapter Four.

Chapter One. Dynamical holism in classical mechanics

Introduction

The term "holism" in general means that in some sense the whole is not the sum of the parts. In the current philosophical discussion, "holism" seems to mean that some properties of a whole system cannot be decomposed into or cannot be determined by the properties of the subsystems. Holism stands opposed to reductionism or atomism, which assume that the properties of the whole supervene on the properties of the parts. Recently, some philosophers have proposed that quantum mechanics implies holism: the state of a composite system cannot be decomposed into the states of the subsystems or the state of a composite system cannot be determined by the states of the subsystems. Some philosophers (such as Scheibe (1995)) argue if the world were holistic, we would have the problem of justifying our scientific practice of applying idealized laws to the systems in the world, i.e., the exportation problem, since idealized laws are supposed to be applied only to isolated or independent systems.

In this chapter I will introduce a new type of holism, namely, dynamical holism. Dynamical holism (which I will argue in the next two chapters is a characteristic of classical mechanics) implies a kind of nonreducibility that makes idealized laws not even *approximate* the descriptions of these dynamically holistic systems. This will raise the exportation problem: how can we be justified in using idealized models to describe dynamically holistic systems? My goal for this chapter is to show that the feature of classical mechanics that I call dynamical holism is real and deserves to be called 'holism'. To do this, I need to show that it meets the criterion associated with holism: some properties of the whole are not determined by the properties of the parts, or more precisely, the dynamical behavior of the whole is not determined by the dynamical behavior of the parts.

The dynamical behavior or motion of a system in classical mechanics is encoded in the Hamiltonian of a system. The motion of a system refers to its time evolution, i.e., the change of the state in time. The Hamiltonian of a system is a function that describes the possible energy configuration of the system. If we know the Hamiltonian of a system, we can write the Hamilton's equation of motion for it and if we can solve the equation, we will get the description of its motion.

However, not all Hamilton's equations are solvable. Whether the Hamilton's equation of motion of a system can be solved is determined by the "separability" of the Hamiltonian. Only when the Hamiltonian is separable, can the equation of motion be solved. The separability of a Hamiltonian is the decomposability of it in terms of the Hamiltonians of its subsystems. For example, the Hamiltonian of a system of N particles is separable if the Hamiltonian of the whole system is the sum of the Hamiltonians of the particles. In contrast, if the Hamiltonian of the whole system is not separable, it is not the sum of the Hamiltonians of the particles or any other subsystems (see Section 1).

When the Hamiltonian of a system is not separable, the equation of motion for the system cannot be solved. This means that the motion of the system cannot be described as the "sum" of the motions of the subsystems. I call such systems dynamically holistic.

This dynamical holism of nonseparable systems has a serious consequence: the motion of a nonseparable system does not become the motion of an idealized separable system even when the difference between the Hamiltonians of the two systems is approaching zero. In philosophical terms, this means that the former cannot be reduced to the latter in the limit of small perturbation, that is, in the limit when the perturbing factors are as small as approaching zero, the descriptions of nonseparable systems do not go over or are not reduced to the descriptions of idealized systems. Since our current views of scientific methods of idealization assume that idealized theories describing idealized systems can be exported to describe real systems because real systems are reducible to idealized systems in the limit of small perturbations, the nonreducibility of nonseparable systems would raise the "new" exportation problem of how to justify our scientific practice of applying idealized theories to real systems if real systems are nonseparable.

The idea that classical mechanics implies dynamical holism may not sound right to some people. They may think that separability is related to the solvability of the equation of motion and thus that nonseparability might be due merely to our not having found a technique to integrate equations with nonseparable Hamiltonians. If so, it would mean that nonseparable systems are actually integrable and thus that a notion of dynamical holism based on nonseparability would not imply nonreducibility in the limit of small perturbation.

However, as I will show in this chapter, nonseparability is not a result of our limited ability in decomposing Hamiltonians and solving Hamilton's equations; rather, it reflects the nature of the dynamical behavior of the nonseparable systems. While separable systems have only regular motion which can be decomposed into the motions of the subsystems, nonseparable systems have both regular and chaotic motion, which cannot be decomposed into the motions of the subsystems. The essential difference between the natures of the dynamical behaviors of separable and nonseparable systems supports the view that separable systems are dynamically atomistic and nonseparable ones are dynamically holistic.

In order to explain the notion of nonseparability and its implication of dynamical holism in classical mechanics, I will introduce some basics of classical mechanics, beginning with Hamiltonian mechanics. Then I will define a notion of dynamical holism in terms of nonseparability. Finally I will answer some possible objections to the notion of dynamical holism, defending the view that the nonseparability of a system's Hamiltonian reflects the nature of the system's dynamical behavior and not our epistemic limitations.

1.1. Basics of Hamiltonian mechanics

In this section I will introduce the basics of Hamiltonian mechanics, such as the concepts of generalized coordinates, generalized momenta and the Hamiltonian of a system. I will explain why canonical transformation plays a crucial role in determining whether a Hamiltonian is separable and thereby whether the motion of the system is decomposable or not.

Hamiltonian mechanics uses the Hamiltonian of a system to describe its motion. A special class of systems comprises Hamiltonian[2] systems, which are the systems that have constant Hamiltonians. Since Hamiltonian systems have simpler mathematical expressions and demonstrate dynamical properties more clearly, my discussion below will focus on Hamiltonian systems.

Hamiltonian mechanics characterizes the state of a Hamiltonian system of n degrees of freedom by n generalized coordinates $q_1, q_2, \ldots q_n$, and n generalized momenta $p_1, p_2, \ldots p_n$. The motion (i.e., the time evolution of the state) of the system $(q_1(t), q_2(t), \ldots q_n(t), p_1(t), p_2(t), \ldots p_n(t))$ is described by the Hamilton's equations:

$$\dot{q} = \frac{\partial H}{\partial p_i}$$

$$\dot{p} = -\frac{\partial H}{\partial q_i} \qquad (i = 1, 2, \ldots n)$$

where $H = H(q_1, q_2, \ldots q_n, p_1, p_2, \ldots p_n)$ is the Hamiltonian of the system and

$$\dot{q} = \frac{dq}{dt}, \dot{p} = \frac{dp}{dt}$$

The motion or time evolution of the system is described by the solutions of Hamilton's equations. The $2n$ variables $q_1, q_2, \ldots q_n, p_1, p_2, \ldots p_n$ are called canonically conjugate variables (whose meaning will be clear after I introduce the concept of canonical transformation).

---

[2] Please note that the term "Hamiltonian" is used to refer both to a property of a system and to a characteristic of the system (having a constant Hamiltonian property). This might be a bit confusing but is the general way of using this term.

The generalized coordinates and momenta $q_1, q_2, \ldots q_n, p_1, p_2, \ldots p_n$ in some cases, e.g., a system of N independent particles, can be understood as the positions and momentums of the particles. But in general cases, they are not.

The n generalized coordinates $q_1, q_2, \ldots q_n$ define the configuration of a system. They may not be Cartesian, i.e., the positions of the subsystems. They can be any set of n variables, which might be any functions of the Cartesian coordinates as long as they can effectively specify the configuration of the system. A special feature of generalized coordinates is that they can reduce the number of the Cartesian coordinates when the systems are subject to holonomic constraints[3]. For example, a system of N particles has $3N$ Cartesian coordinates, namely, $x_1, x_2, \ldots x_{3N}$. If the system is subject to k holonomic constraints, that is, constraints that can be described by k relations of the coordinates in the form

$$G_s(x_1, x_2, \ldots x_{3N}) = 0 \qquad (s = 1, 2, \ldots k)$$

the number of generalized coordinates can be reduced to $3N-k$ by transforming the coordinates from the Cartesian $x_1, x_2, \ldots x_{3N}$ to the generalized coordinates $q_1, q_2, \ldots q_{3N}$ such that the constraints become

$$q_{3N-k+s} = 0 \qquad (s = 1, 2, \ldots k)$$

---

[3] A holonomic constraint is one that can be expressed by equations satisfied by the Cartesian coordinates: $G_l(x_1, x_2, \ldots x_{3N}) = 0$

The remaining nonzero $n = 3N\text{-}k$ generalized coordinates are

$$q_m = q_m(x_1, x_2, \dots x_{3N}) \qquad (m = 1, 2, \dots 3N\text{-}k)$$

Therefore the configuration of the system depends only on the n independent generalized coordinates $q_1, \dots q_n$ ($n = 3N - k$) (Corben, 1960, p.10) and the system is said to have n degrees of freedom. This indicates that the n generalized coordinates $q_1, q_2, \dots q_n$ must be independent of each other and are sufficient to describe the configuration of the system.

The n generalized momenta $p_1, p_2, \dots p_n$, are derived from the Lagrangian of the system. The Lagrangian is a function $L(q_1, q_2, \dots q_n, \dot{q}_1, \dot{q}_2, \dots \dot{q}_n)$ where the $\dot{q}_1, \dot{q}_2 \dots \dot{q}_n$ are the generalized velocities. For example, the Lagrangian for a system of N interacting particles with n degrees of freedom is:

$$L = T - V = \sum_{i=1}^{n} \frac{1}{2} m_i \left(\frac{dq_i}{dt}\right)^2 - V(q_1, q_2, \dots q_n)$$

where T is the kinetic energy and V is the potential energy of the system. Generalized momenta are defined as

$$p_i = \frac{\partial L}{\partial \dot{q}_i}$$

Only in Cartesian coordinates where

$$p_i = m_i \dot{q}_i$$

are the p_i the momenta of the particles. In general, generalized momenta are independent variables, which are not necessarily related to the generalized velocities in the Lagrangian function. Rather, they have completely different geometric properties than velocities as I will explain below (Tabor, 1989, p.47).

The Hamiltonian of a Hamiltonian system is derived from its Lagrangian

$$H(\boldsymbol{p},\boldsymbol{q}) = \sum\nolimits_{i=1}^{n} p_i \dot{q}_i - L(\boldsymbol{q},\dot{\boldsymbol{q}})$$

by replacing the $\dot{\boldsymbol{q}}$ by $\boldsymbol{p}$ (ibid., p.49) (here the bold $\boldsymbol{p}$, $\boldsymbol{q}$ represent the vectors $\boldsymbol{p} = (p_1, p_2, \ldots p_n)$ and $\boldsymbol{q} = (q_1, q_2, \ldots q_n)$). In the case of a system of N interacting particles, the Hamiltonian is

$$H(\boldsymbol{p},\boldsymbol{q}) = \sum\nolimits_{i=1}^{n} \left(\frac{1}{2m_i}\right) p_i^2 + V(q_1, q_2, \ldots q_n) = T + V$$

In this case, the Hamiltonian is the total energy of the system and is a constant of motion.[4] Since the motion of a system is determined by the Hamilton's equations together with its Hamiltonian, the Hamiltonian of a system in effect determines all of the possible motions the system can perform in phase space. As the physicist Herbert Goldstein put it, "in a literal sense, the Hamiltonian is the generator of the system motion with time." (Goldstein, 1950, p.259).

---

[4] The conditions for the Hamiltonian to be the total energy are: the constraints do not depend explicitly on time and the potential energy is a function of the position only. If the Hamiltonian is the total energy and if the system is conservative, i.e., its total energy is invariant, the Hamiltonian

It might seem that the description of the motion of a composite system, for example, a system of *N* particles of 3*N* degrees of freedom, is automatically the "sum" of the motions of the subsystems (represented by a pair of generalized coordinate and momentum, e.g., p_n, q_n) since it is represented by the position and momentum variables of the subsystems (i.e., $p_1, q_1, p_2, q_2, \ldots p_n, q_n$). However, when the subsystems are interacting with each other, the Hamiltonian of the system in terms of the generalized coordinates and momenta would have terms that contain the variables of different subsystems (such as $H_{\text{int}}(q_i, q_j)$, for example, in the case of N-body problem, $H_{\text{int}} = \sum_{1\le i<j\le N} \frac{Gm_i m_j}{\|q_i - q_j\|}$ where *G* is the universal gravitational constant, and m_i, m_j are the masses of the *i*th and *j*th bodies. Because of these so called coupled or interaction terms, each of the Hamilton's equations of motion involves more than one pair of variables. For example,

$$\dot{q}_i = \frac{\partial[H_1(p_1,q_1) + H_2(p_2,q_2) + \ldots H_n(p_n,q_n) + \sum_{i=1,i\neq j}^{n} H_{\text{int}}(p_i,q_i,q_j)]}{\partial p_i},$$

$$\dot{p}_i = -\frac{\partial[H_1(p_1,q_1) + H_2(p_2,q_2) + \ldots H_n(p_n,q_n) + \sum_{i=1,i\neq j}^{n} H_{\text{int}}(p_i,q_i,q_j)]}{\partial q_i}$$

---

is the constant integral of energy. (Groesberg, p. 284, Goldstein, p.220). As Groesberg notes, the Hamiltonian may be a constant but not the total energy when the Hamiltonian does not depend on time but the constraints depend on time. So nonconservative systems may have constant Hamiltonians (Groesberg, p.284).

which makes these variables no longer independent of each other and thus the subsystems do not have independent motions. This makes it impossible to describe the motion of the whole system in terms of the *independent* motions of the subsystems and thus in this sense the motion of the whole system is not the sum of, or cannot be derived from summing up the independent motions of the subsystems.

We may ignore the interaction terms in the Hamiltonian and solve the Hamilton's equation of motion. The solution would describe the independent motions of the simplified or noninteracting subsystems. But just as the Hamiltonian is not the sum of the Hamiltonians of the noninteracting subsystems, the motion of the whole system is also not the sum of the motions of the noninteracting subsystems.

However, not all systems that have Hamiltonians with interaction terms fail to have solutions. The Hamiltonian of a system can be transformed to an equivalent Hamiltonian in terms of a new set of generalized coordinates and momenta through a mathematical technique, namely canonical transformation, because of the invariance of the Hamiltonian under canonical transformation (see Appendix). That is, by changing the generalized coordinates and momenta by canonical transformation, we will obtain a Hamiltonian described in terms of new generalized coordinates and momenta. The invariance of the Hamiltonian under canonical transformation means that the "new" Hamiltonian describes the same dynamical system as the original one. They are different representations of the same system.

The invariance of Hamiltonians makes it possible for us to use canonical transformation to separate the Hamiltonians of some systems and solve the

Hamilton's equations. This is because we might be able to use canonical transformation to "eliminate" the interaction terms in a Hamiltonian and thus the "transformed" Hamiltonian would not have interaction terms. This makes it possible to solve the equation of motion of the system and describe its motion in terms of the transformed variables which would be independent of each other. So the motion of the whole system would then be the sum of the independent motions of the "transformed subsystems" (represented by a pair of the transformed generalized coordinate and momentum) although it is not the sum of the motions of the original subsystems or the simplified original subsystems. For example, the Hamiltonian of a system composed of a pair of harmonic oscillators coupled by a spring is

$$H = \frac{1}{2}({p_1}^2 + {p_2}^2 + {q_1}^2 + {q_2}^2) + \frac{1}{2}q_1 q_2$$

where the term $\frac{1}{2}q_1 q_2$ is the interaction term, which would make each pair of Hamilton's equations

$$\dot{q}_i = \frac{\partial H}{\partial p_i} \quad \dot{p}_i = -\frac{\partial H}{\partial q_i}$$

involve more than one pair of variables. So the Hamiltonian is not the sum of the Hamiltonians of the subsystem

$$H \neq \frac{1}{2}({p_1}^2 + {p_2}^2 + {q_1}^2 + {q_2}^2)$$

However, if we apply the canonical transformation:

$$Q_1 = \frac{q_1 + q_2}{\sqrt{2}}; Q_2 = \frac{q_1 - q_2}{\sqrt{2}}; P_1 = p_1; P_2 = p_2$$

the Hamiltonian can be transformed to

$$H = \frac{P_1^2}{2} + \frac{{Q_1}^2}{4} + \frac{{P_2}^2}{2} + \frac{3{Q_2}^2}{4}$$

which is equivalent to the original Hamiltonian but it is the sum of the Hamiltonians of the subsystems represented by (P_1, Q_1) and (P_2, Q_2) (This example is from Kronz, 1998, p.58).

The Hamilton's equations in terms of the new variables P_i, Q_i can then be solved, which means that the motion of the whole system can be seen as the sum of the motions of the subsystems. The solution can in principle be inversely transformed back to the solution in terms of the old variables p_i's, q_i's (Tabor, 1989, p.58). These solutions in terms of different coordinate systems can be seen as different representations of the motion of the same system. But if the interaction terms in the Hamiltonian of a system cannot be eliminated by canonical transformation, the Hamiltonian cannot be the sum of the Hamiltonians of any subsystems (i.e. in terms of the *old* or the *new* generalized coordinates and momenta) and the Hamilton's equations cannot be solved. Therefore, the motion of the whole system is not the summation of the independent motions of noninteracting subsystems.

Therefore whether the Hamilton's equation of a system can be solved and thus whether its motion can be the sum of the independent motions of the subsystems represented by the pairs of independent generalized coordinates and momenta is determined by whether the interaction terms in the Hamiltonian can be removed by canonical transformation. This important property is called the separability/nonseparability of the Hamiltonian.

1.2. The nonseparability of Hamiltonians and the nonintegrability of Hamiltonian systems

As we have seen from the last section, some Hamiltonians either do not contain interaction terms or have interaction terms which can be removed by canonical transformations, whereas other Hamiltonians contain "nonremovable" interaction terms. This property – the separability of the Hamiltonian – determines not only whether the Hamiltonian can be the sum of those of the subsystems, but also whether the motion of the whole system can be the sum of the motions of the subsystems. That is, the separability of the Hamiltonian determines whether the system is dynamically holistic.

To be precise, I will introduce the formal or technical definition of separability of a Hamiltonian here. The separability of a Hamiltonian in Hamiltonian mechanics is defined as: The Hamiltonian of a system of n degrees of freedom, $H\ (\boldsymbol{p},\ \boldsymbol{q}) = \sum_{i=1}^{n} H_i(p_i,q_i) + H_{\text{int}}(p_1,p_2,\dots p_n,q_1,q_2,\dots q_n)$ is separable if, and only if, there exist canonical transformations such that it can be written as the sum of the Hamiltonians, $H_i^{'}(P_i,Q_i)$, in the form

$$H'(\boldsymbol{P},\boldsymbol{Q}) = \sum_{i=1}^{n} H_i^{'}(P_i,Q_i)\ .$$

So the interaction term(s) represented by $H_{\text{int}}(p_1,p_2,\dots p_n,q_1,q_2,\dots q_n)$ in the separable Hamiltonian *H*(***p*, *q***), is absorbed into the Hamiltonians at each degree of freedom in terms of the "new" generalized coordinates and momenta, $H_i^{'}(P_i,Q_i)$, through canonical transformation. Canonical transformation can effectively eliminate coupled or interaction terms, that is, the terms with more than one pair of variables, in separable Hamiltonians). However, canonical transformation cannot remove the coupled terms in nonseparable Hamiltonians. For nonseparable Hamiltonians, no canonical transformation can make the above equation hold.

Although for many other sets of generalized coordinates and momenta on other canonical transformations, the Hamiltonian of a separable system is not decomposable, i.e., cannot be written as the sum of the Hamiltonians at each degree of freedom represented by each pair of the generalized coordinates and momenta, the Hamiltonian of a separable system must be decomposable in terms of at least one set of the generalized coordinates and momenta. However, there might be different canonical transformations that can make the Hamiltonian of a separable system decomposable in terms of different sets of generalized coordinates and momenta. In contrast, a nonseparable Hamiltonian is not decomposable in terms of any set of generalized coordinates and momenta.

We may see each pair of the generalized coordinates and momenta, q_i, p_i (or Q_i, P_i) describing the motion at each degree of freedom of a system of n degrees of freedom as describing the motion of a subsystem of it. Since generalized coordinates and momenta include both the original ones described by q_i, p_i and the transformed ones represented by Q_i, P_i, we can distinguish between these two types of generalized subsystems: we call the former "ordinary" subsystems and the latter "theoretical" subsystems. They are so called because the former are associated with the ordinary objects that constitute a composite system more directly than the latter. For example, in the equations of the many-body problem on p.16, each pair of p_i, q_i representing an ordinary subsystem, describes the motion of one body of the system at one of its three degrees of freedom. In contrast, when we use canonical transformation to transform the variables p_i, q_i to new variables P_i, Q_i, the new variables, representing theoretical subsystems, are functions of p_i, q_i and thus may not be directly related to the bodies that constitute the system as the ordinary subsystems are related. In this sense we may say that the theoretical subsystems are more abstract than the ordinary subsystems. Another difference between ordinary and theoretical subsystems is that for some of the systems consisting of *interacting* bodies, we may be able to write the Hamiltonian of the system as the sum of the

Hamiltonians of a set of its theoretical subsystems but not as the sum of those of its ordinary subsystems. When the Hamiltonian of a composite system is the sum of the Hamiltonians of its generalized subsystems, those subsystems are called independent, whether they are theoretical subsystems or ordinary subsystems. When the Hamiltonian is not a sum of the Hamiltonians of the subsystems, the subsystems are called dependent.

For non-integrable systems, we may regard the failure to be able to write the Hamiltonian of the system as a sum of the Hamiltonians of a set of generalized subsystems in either of two ways. First, we might say that since the generalized subsystems are not independent, they do not in fact have well-defined Hamiltonians. Alternatively, even if we ignore the interactions among these dependent generalized subsystems so as to assign Hamiltonians to these non-interacting "artificial" subsystems, the Hamiltonian of the system still cannot be written as their sum because there would be no term representing the interactions.

Another way of defining the notion of separability (which is equivalent to the above definition) is in terms of action-angle variables. Since this definition is easier to use in practice and will be used in the later discussions, I will also introduce it here (see Helleman, 1980, p.170): a Hamiltonian is separable if and only if it can be transformed into action-angle variables $(\boldsymbol{I}, \boldsymbol{\theta})$, where $\boldsymbol{I} = \boldsymbol{I}(p_1, p_2, \ldots p_n, q_1, q_2, \ldots q_n)$ is the action vector and $\boldsymbol{\theta} = \boldsymbol{\theta}\, (p_1, p_2, \ldots p_n, q_1, q_2, \ldots q_n)$ is the angle vector which increases by 2π after each complete period of motion[5], so that the transformed Hamiltonian is a function of the action variables I_n only, that is,

---

[5] Note that Hamiltonian systems have periodic motion.

$$H'(\boldsymbol{I}) = \sum_{i=1}^{n} I_n$$

The notion of separability is closely related to the notion of integrability, i.e., the solvability of equation of motion. The equations of motion involving separable Hamiltonians are solvable. This can be clearly seen from the Hamilton's equations for a separable system in terms of action-angle variables:

$$\dot{I}_i = -\frac{\partial H'}{\partial \theta_i} = 0$$

$$\dot{\theta}_i = \frac{\partial H'}{\partial I_i} = \omega_i \qquad (i = 1, 2, \ldots n)$$

which have the solutions:

$$I_i = const.$$

$$\theta_i = \omega_i t + const.$$

A system whose equations of motion are solvable is called integrable. The integrability of Hamiltonian systems is defined as the following: A Hamiltonian system of n degrees of freedom is completely integrable,[6] if and only if, there exist n

---

[6] Note the fact that some Hamiltonian systems that may not be completely integrable, may be integrable in some sense, and that some nonHamiltonian, i.e., dissipative systems can be integrable requires a more general notion of integrability as Flaschka et al. argue (Flaschka et al. 1991 p.74-

independent integrals (constants of motion, $f_i(\boldsymbol{p},\boldsymbol{q})$ = *const.* (i= *1, 2, ...n*)) in involution, that is, they all commute with each other. Their Poisson brackets are 0, that is, ([g_i, f_i] = *0*). This means that the *n* independent integrals form a complete set of new momenta in some transformed coordinate system (Lichtenberg and Lieberman, 1983, p.23). In general, most systems are nonintegrable, i.e., since they do not have a sufficient number of independent integrals in involution, they cannot be solved. But as we saw above, a system with separable Hamiltonian is integrable since the *n* H_i or I_i can be taken as *n* independent integrals in involution (Arnold, 1983, p.125).

In contrast, since the Hamiltonian of a nonseparable system is not the sum of the Hamiltonians of its generalized subsystems, its equations of motion cannot be the set of the independent equations of the generalized subsystems either. We can see this from its equations of motion (in terms of action-angle variables):

$$\dot{I}_i = -\frac{\partial H'(I_1, I_2, \cdots I_n, \theta_1, \theta_2, \cdots \theta_n)}{\partial \theta_i}$$

$$\dot{\theta}_i = \frac{\partial H'(I_1, I_2, \cdots I_n, \theta_1, \theta_2, \cdots \theta_n)}{\partial I_i} \qquad (\, i = 1,\, 2,\, \ldots n)$$

where the Hamiltonian is nonseparable, that is, it has coupling terms which contain not only $\boldsymbol{I}_i, \theta_i$ but also $\boldsymbol{I}_j, \theta_j$ ($I \neq j$) etc. This makes the equations of nonseparable systems not solvable, that is, we cannot write the solutions in a closed functional

---

75). But my discussion will focus on Hamiltonian systems and their complete integrability which is closely related to the separability of Hamiltonian.

form. This is because not all the I_i, θ_i can be characterized as independent functions; some of them have to be dependent on some other variables.

As Kronz notes (Kronz, 1998, p.58), the separability of a Hamiltonian is logically equivalent to the integrability of the corresponding system. As we saw above, since separable systems always have *n* independent integrals in involution they are integrable. Conversely, when there are *n* integrals for a Hamiltonian system with *n* degrees of freedom, each of the integrals can be used to separate a pair of conjugate variables and reduce one degree of freedom of the system. So *n* integrals can reduce the system to *n* problems of one degree of freedom (which are solvable) (see Whittaker, 1952, pp. 313-314, Lichtenberg and Lieberman, 1983, p.23). Noting the equivalence of the separability of a Hamiltonian and the integrability of the corresponding system, Helleman defines integrability in terms of the possibility of transforming the system into action-angle variables (Helleman, 1980, pp.170-171). So the problem of integrability becomes the problem of finding a canonical transformation that transforms a system into action-angle variables, where the transformed Hamiltonian is explicitly separable.

To avoid confusions about the different notions I have introduced above, I will use a diagram to show the relations among them:

Separability ⇔ decomposability of Hamiltonian ⇔ integrability ⇔ solvability

Where the symbol "⇔" means "entails and is entailed by"

So far I have shown that the separability of a system determines the solvability of its equation of motion. Since the solution of a separable system can be seen as the "sum" of the solutions of the generalized subsystems ($I_i = const.$, $\theta_i = \omega_i t + const.$ ($i=1, 2...n$)), we can see the motion of a separable system as the sum of the motions of the independent generalized subsystems. In contrast, the motion of a nonseparable system is not the sum of the independent motions of its "artificial" independent generalized subsystems. Since the motions of its generalized subsystems are dependent on each other, the motion of a nonseparable system cannot be determined by the motions of the generalized subsystems. So the separability of a system determines the decomposability of its motion into the motions of its subsystems.

1.3. The notion of dynamical holism defined in terms of nonseparability

As I have shown above, in classical mechanics, the Hamiltonian (or motion) of a separable system is the "sum" of the Hamiltonians (or motions) of its independent generalized subsystems whereas nonseparable systems do not have independent generalized subsystems and thus the Hamiltonian (or motion) of a nonseparable system is not the "sum" of the Hamiltonians (or motions) of its generalized subsystems.

Since nonseparable systems have the feature that the dynamical behavior of the whole cannot be determined by that of the subsystems, we can call nonseparable systems dynamically holistic systems. Since it is the nonseparability of the Hamiltonian of a system which determines the nondecomposability of its motion, we may define the notion of dynamical holism as the following: *a Hamiltonian system is dynamically holistic if and only if its Hamiltonian is nonseparable.*[7] The notion of dynamical holism means that the Hamiltonian and motion of a whole system is not the sum of or is not determined by that of the subsystems. Conversely, we can define the notion of dynamical atomism as meaning that the Hamiltonian and motion of a whole system is the sum of that of the subsystems.

So all nonseparable Hamiltonian systems are dynamically holistic whereas all separable systems are dynamically atomistic.

---

[7] This definition may be extended to include quantum dynamical holism. In quantum mechanics, nonseparable Hamiltonians also produce nonseparable evolution, but in the sense that the density operator of the subsystems cannot be characterized independently of the whole (see Kronz and Tiehen, pp.343-344). However, the nonseparable Hamiltonian in classical mechanics implies chaotic motion of the system, whereas in quantum mechanics it produces quantum chaos only in continuous open systems. (see Kronz, 1998).

1.4. Some possible objections to the notion of dynamical holism

(a) Objection: dynamical holism is merely an epistemic problem

Some may worry that since separability is closely related to our solvability of the equation of motion, it may be an indication of our epistemic capability rather than the ontological decomposability of the motion of a system. If nonseparable systems actually could be solved in terms of some generalized coordinates and momenta that are just unknown to us, their motion would be actually decomposable into the motions of the generalized subsystems. To this worry, Tabor has a very clear answer: "our ability to integrate an equation is more than a mathematical nicety but touches right at the heart of the deepest properties of a system" (Tabor p. 3). Below, I will show that the separability (or nonseparability) of a system is not determined by our ability to solve its equation of motion but by the nature of its dynamical behavior, that is, separable systems have regular motion which can be decomposed into the motions of a set of generalized subsystems, whereas nonseparable systems have both regular and chaotic motions which cannot be decomposed into the motions of its generalized subsystems.

The difference between the dynamical behavior of separable and nonseparable systems can be clearly seen from their different phase portraits. A phase portrait represents the motion or dynamical behavior of a system in a phase space defined by the *2n* conjugate variables of the system (of *n* degrees of freedom).

The state of the system at a given time corresponds to a point in phase space and the motion of the system in time corresponds to a trajectory in phase space. Different initial conditions correspond to different trajectories and all the trajectories of the system, which constitute a phase portrait, provide a complete description of the dynamical behavior of the system. For example, the phase portrait of a simple harmonic oscillator with frequency ω, described by the equations:

$$\dot{x} = y$$

$$\dot{y} = -\omega^2 x$$

and whose Hamiltonian is

$$H = \frac{1}{2}(y^2 + \omega^2 x^2)$$

comprises a set of concentric ellipses (see Figure 1). The major and minor axes of the ellipses are determined by the Hamiltonian. Thus, the Hamiltonian provides a definite geometrical constraint on the trajectories of a Hamiltonian system: it confines its motion to these closed curves. Note that the time dimension is suppressed in the phase portrait. If we try to represent the motion of the above system in a space adding a time dimension perpendicular to the *x, y* plane, we will get spiral curves.

The phase portrait of a *separable* Hamiltonian system of *n* degrees of freedom is composed of trajectories that are confined to some *n*-dimensional torus

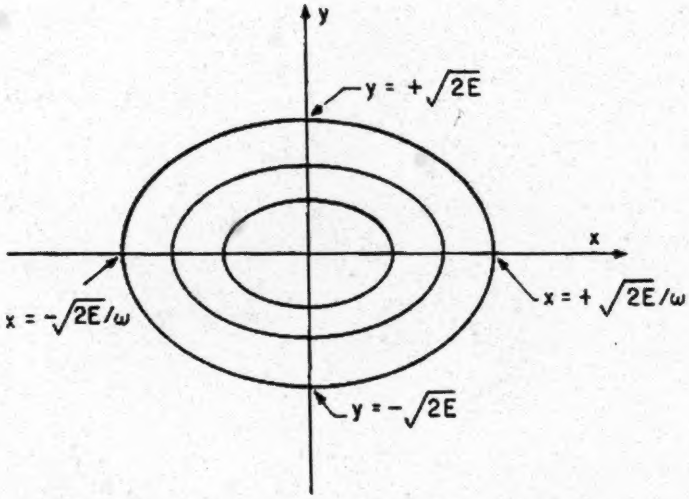
y
y = +$\sqrt{2E}$
x
x = $-\sqrt{2E}/\omega$
x = $+\sqrt{2E}/\omega$
y = $-\sqrt{2E}$

Figure 1.2 Phase curves for the simple harmonic oscillator (1.1.4).

Figure 1

From M. Tabor, 1989, P.14, figure 1.2

and the *2n* dimensional phase space of the system is filled with *n* dimensional tori. Each solution of the Hamilton's equations of the system corresponding to a certain initial condition maps out a smooth *n* dimensional surface in the *2n* dimensional phase space. The important property of this surface (a torus) is that it is the direct product[8] of *n* independent closed paths or curves (in the subspaces), each describing an independent periodic motion with the action variable

$$I_i = \frac{1}{2\pi} \int_{r_i} \sum_{m=1}^{n} p_m dq_m$$

and the angle variable

$$\theta_i = \omega_i t + const.$$

That is, we can define *n* topologically independent closed paths on the torus where none of them can be deformed continuously into each other or shrunk to zero (see Figure 2). This means that the phase portrait of a separable system, i.e., the tori, can be decomposed into (the product of) the independent trajectories of the independent generalized subsystems. So the motion of a separable system of *n* degrees of freedom is equivalent to *n* independent pendula or *n* harmonic oscillators of one degree of freedom. (Helleman, p.171).

The tori of a separable system are of two kinds, which change very differently under perturbation (i.e., a small change in the Hamiltonian of the system),

---

[8] We may think of the *n* independent paths in the subspaces as the projections of the torus at each subspace represented by (p_i, q_i) and the direct product of these paths form the torus.

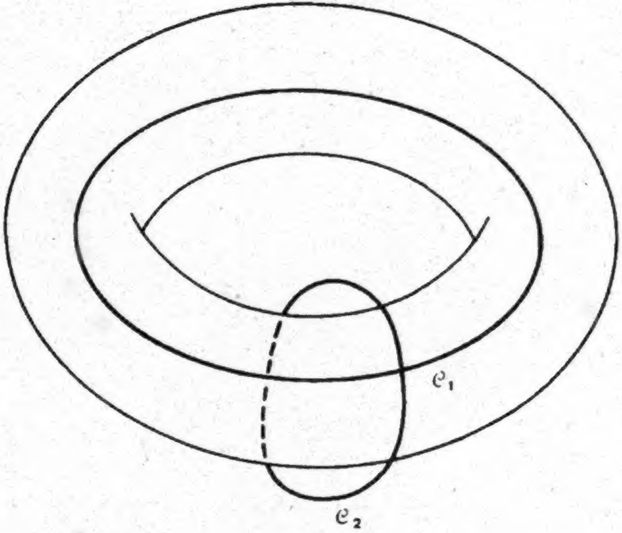
$\mathscr{C}_1$
$\mathscr{C}_2$

Figure 2.2 The two topologically independent curves $\mathscr{C}_1$ and $\mathscr{C}_2$ on a 2-torus.

Figure 2

From M. Tabor, 1989, P.72, figure 2.2

a fact which accounts for the topological difference of the phase portrait of a nonseparable system from a separable one. If the frequencies ω_i of a torus are all rationally related, i.e.,

$$\sum_{i=1}^{n} k_i \cdot \omega_i = 0$$

where the k_i's are integers, the motion on the torus will be a closed orbit, that is, the motion will repeat itself. Conversely, if the frequencies are not rationally related, the motion will never repeat itself and a single orbit covers the torus "uniformly" (see Tabor, pp.73-74, fn). Such orbits are called quasi-periodic. If the frequencies are functionally independent, that is, (Arnold, 1988, p.183),

$$\det\left|\frac{\partial \omega_i(I)}{\partial I_j}\right| = \left|\frac{\partial^2 H(I)}{\partial I_i \partial I_j}\right| \neq 0$$

the system is called nondegenerate (Tabor, 1989, p.78). This means that the frequencies vary from torus to torus. Although there are infinitely many rational numbers, they are infinitely outnumbered by the irrationals (ibid., p.78). Thus, a nondegenerate system has an infinite number of tori with closed orbits and infinitely many more tori with quasi-periodic orbits.

In contrast, the phase space of a nonseparable system is not filled with tori but is much more complicated. A nonseparable system does not have n independent integrals in involution to make it possible to decompose its motion into n

independent periodic motions. If we try to transform the description of a nonseparable system into a description in terms of action-angle variables in order to separate the subsystems, we will find that the Birkhoff series, which we get when we write each of the old variables as a power series of the action and angle variables, diverges (See Helleman, 1980, p.170). This is because the term $\sum_{i=1}^{n} k_i \cdot \omega_i$ appears in the denominators of the terms in the series.[9] When $\sum_{i=1}^{n} k_i \cdot \omega_i = 0$ or is arbitrarily small, the terms with such denominators will blow up or be arbitrarily large.[10] Helleman attributes this divergence to the presence of a singularity in one of the integrals $I_i(q_1, \cdots q_n, p_1, \cdots p_n)$ (with $i = 1, \ldots, n$) and concludes that "there appears to be infinite set of such singularities … in phase space…. This makes it hard to explicitly display such integrals" (ibid., p.172). The singularities in the integrals make it impossible to decompose the surfaces (or the phase portrait, which is composed of the phase space trajectories) of a nonseparable system of say, *n* degrees of freedom, into the product of *n* independent trajectories in the subspaces, each of them representing one integral I_i. This is why the Hamiltonian of a nonseparable system cannot be written as the sum of the Hamiltonians of its "artificial" independent generalized subsystems and its solution is not the "sum" of the solutions of these subsystems.

What does the phase portrait of a nonseparable system look like then? Would all the tori disappear in the phase space of a nonseparable system? The famous KAM theorem shows that if the nonseparable system is near-integrable, that is, it can be seen as an integrable system plus some small nonintegrable perturbation, and if the unperturbed integrable system is nondegenerate, which means that most of the tori of the unperturbed system are covered with quasi-periodic orbits, most of the tori of the unperturbed system, the ones covered with quasi-periodic motion or orbits, will not be destroyed but only slightly deformed. These tori are called KAM tori. The tori with rational frequencies, which give rise to divergent Birkhoff series of nonseparable systems, will be "destroyed" under perturbation and the structure of these regions becomes extremely complicated. (see Tabor, 1989, pp.139-141). In Lichtenberg and Lieberman's words, "[u]nder perturbation, some solutions associated with particular resonances change their topology" and the change of topology leads to the formation of chaotic regions. (Lichtenberg and Lieberman, 1983, pp.63-64). That is, when the Birkhoff series diverges, the motion (in the regions where the Birkhoff series is divergent) is chaotic (Helleman, 1980, p.175). The trajectories in the chaotic regions are not confined to a torus but wander over larger regions on the energy shell and form chaotic regions, where the motion is sensitive to initial conditions and appears random (Tabor, 1989, p.34). The change of topology means that the topology of the phase portrait of the perturbed system, i.e., the nonseparable system, is so different from that of the unperturbed, i.e., separable system, that no canonical transformation can transform one into the other. The bigger the nonseparable perturbation, the bigger the chaotic regions. Therefore, the phase space portraits of nonseparable systems have both regular regions, where most orbits are confined to KAM tori, and small irregular or chaotic regions between

---

[9] Note that all Hamiltonian systems of one degree of freedom are integrable (Helleman, p.170) and resonances occur when there is more than one degree of freedom (between the fundamental frequencies) (Lichtenberg and Lieberman, p.64).

[10] Lichtenberg and Lieberman note, "this phenomenon represents a physical as well as a mathematical difficulty. It arises from real resonances that, …change the topology of the phase space trajectories." (Lichtenberg and Lieberman, p.77)

KAM tori with wandering orbits. However, there are also periodic orbits in the chaotic regions and some of them are closed orbits. This means that the phase space of a nonseparable system is filled with a hierarchy of regular and chaotic regions, "inside the regular (smooth) regions are chaotic regions, inside which are regular regions, etc." (Helleman, 1980, p.182).

The topological difference between the phase portrait of integrable and nonintegrable systems implies that the behavior of integrable systems is *qualitatively* different from the behavior of nonintegrable systems: the integrable systems have only regular motion whereas nonintegrable systems have both regular and chaotic motion. The qualitative behavior of a dynamical system is represented by the topology of its phase portrait. The behavior of a system is qualitatively equivalent to the behavior of another system if it is topologically equivalent to the latter, that is, there exists a homeomorphism that continuously maps trajectories of the former system to those of the latter system (see Arnold, 1983, Ch. 3; also Rueger, 2000, p.319 and p.303). Since there is no smooth deformation that can transform the chaotic trajectories of nonseparable systems to the regular trajectories of separable systems, the behavior of nonseparable systems is qualitatively different from the behavior of separable systems. The qualitative difference between regular and chaotic motion can be understood from the fact that the chaotic motion of a nonseparable system cannot be described analytically or by convergent perturbation series because, as I noted above, chaotic trajectories are associated with the resonances among the degrees of freedom, which generate divergent perturbation series. As Lichtenberg and Lieberman say, "Perturbation theory fails to describe the complexity of this chaotic motion; formally, the series diverge." (Lichtenberg and Lieberman, 1983, p.63) Also, as I quoted Helleman above, the divergence of the perturbation series means singularities in one of the integrals, which makes it impossible to describe the motion analytically as smooth functions of time. The significant topological change in the chaotic regions of phase space implies that "perturbation methods fail to describe even qualitatively the chaotic motion in regions for which the main resonances overlap." (Ibid., p.64)

We have seen that separable and nonseparable systems have topologically different phase portraits, which indicates that they have qualitatively different behaviors. Besides using phase portraits to describe the motion of a system, classical mechanics also uses the so-called surface of section to characterize the motion of a system. Surface of section enables us to describe the motion of a system of *n* degrees of freedom in a two-dimensional surface so that we can show the behavior of the system by a two-dimensional picture. A surface of section is constructed by recording the intersections of the trajectories of a system in phase space with a chosen surface, for example, the plane $p_i, q_i,$ or p_x, x, where the trajectories pierce the chosen plane in a certain direction, say, from bottom to top or from inside to outside of the plane. The motion of a system can be described by the points in the plane, that is, its surface of section. The surface of section of an integrable system either has finite points when the orbits on the tori are closed, or has infinite numbers of points which fill up smooth curves when the orbits on the tori are quasi-periodic. By contrast, the surface of section of a nonseparable system has an indefinite number of random points which may fill up areas of the surface of section when the orbits are in the chaotic regions. The surfaces of section of Henon-Heiles clearly show this (see Tabor, 1989, p.122, Helleman, 1980, p.174, Lichtenberg and Lieberman, 1983, p.49). So the surfaces of section of separable and nonseparable systems show that the motion of nonseparable systems is chaotic, which is essentially different from the regular motion of separable systems. This indicates that the differences between the behaviors of separable and nonseparable systems are real and not due to the particular way we describe the systems.

The difference between the regular motion of separable systems and the chaotic motion of nonseparable systems also can be described in computational ways; this fact also supports the view that the difference is real. One such way employs Lyapunov exponents which describe the mean rate of exponential divergence of neighboring trajectories. For regular motion, the Lyapunov exponents are zero whereas for chaotic motion they are positive. This indicates that the neighboring trajectories of chaotic motion diverge exponentially whereas those of regular motion diverge only linearly (Tabor, 1989, p.148). This implies that chaotic motion is very sensitive to initial conditions. A small change in initial conditions can lead to vastly different subsequent behavior. In contrast, the behavior of a separable system would not change vastly when its initial conditions change. This is another essential difference between the dynamical properties of nonseparable systems and those of the separable systems.

From the above technical discussion of the differences between the dynamical behaviors of separable and nonseparable systems, we can conclude that the dynamical nondecomposibility of nonseparable systems is not due to the fact that we have limited epistemic ability to separate Hamiltonians and to solve the equations but instead to the very nature of the dynamical behavior of nonseparable systems.

(b) Objection: dynamical properties supervene on states

We should note that the holism demonstrated by nonseparable systems in classical mechanics is dynamical rather than state holism: state holism refers to the nonreducibility of states of systems to the "sums" of the states of their subsystems whereas dynamical holism refers to the nonreducibility of the dynamical properties of systems to the "sums" of the dynamical properties of the subsystems. The state of a system at a given time is defined as the values of its position and momentum at that instant of time whereas the dynamics (or motion) of the system refers to the time evolution of the state of the system. The dynamical holism in classical mechanics means that the motion, i.e., the change of the state in time, of a whole system is not the sum of the motions of the subsystems although the state of the whole system is the sum of the states of the subsystems.

This may raise a possible objection to the notion of dynamical holism: if we can see the motion of a system as the collection of the states of the system at each moment of time and if the states of the system are the sum of the states of the subsystems, then the motion of the system can be seen as the collection of the "sum" of the states of the subsystems at each moment of time.

To this objection, I will respond that the notion of dynamical holism is different from the notion of state holism and whether the motion of a system is decomposable cannot be determined by the reducibility of the states of the system. The notion of state or system holism is generally defined as: the state (or all physical properties) of a composite system cannot be reduced to or does not supervene on the nonrelational states of its parts (or component subsystems) together with the spatial relations between the parts (see for example, Maudlin, 1998, and Healey, 1999). It seems to be a general view that entangled states in quantum mechanics demonstrate state holism: the entangled state of a composite system does not supervene on the state of its subsystems[11]. It is also generally believed that there are no cases of state holism in classical mechanics: the state of a classical system can always be reduced to the sum of the states of its subsystems. However, although the classical state of a system (with separable Hamiltonian or not) at each instant of time can be reduced to the sum of the states of its subsystems, its motion is not the sum of the motions of its subsystems when its Hamiltonian is not separable. As I have shown in Section 1.4 (a), the trajectories of a nonseparable system in the phase space cannot be decomposed into the product of the trajectories in the subspaces although each point of the trajectories, i.e., the state of it at each instance of time, is a collection of (and thus can be decomposed into) the values of the coordinates or generalized coordinates and momenta. This is because smooth trajectories cannot be drawn in the subspaces through some of the points (if they are singular) although they have definite values. That is, although we can use a point in phase space to describe the state of the system at any instant of time (thus each coordinate has a definite value), it is not possible to draw smooth trajectories for each degree of freedom as a function of time. This is because some of the integrals I_k have singularities (see p.33), which means that they are not *continuous single-valued* functions and thus cannot be drawn as smooth paths or trajectories. We may use the (insufficiently many) smooth integrals (if they exist) to separate some of the variables-- just like holonomic constraints, each integral can separate one pair of the conjugate variables from the other variables and thus separate one degree of freedom of the system from the others. But since there are not as many integrals as the number of degrees of freedom, there must be some variables that are not separable and cannot be described as independent smooth functions.

---

[11] Different interpretations have different explanations for the state nonreducibility. For the interpretations that hold that the subsystems of a composite system in an entangled state do not have their own states (such as the no-states ray view (so-called by Maudlin)), the state nonreducibility is due to the fact that there are no states to reduce to. For the interpretations that think the subsystems have their own states (such as the statistical operator view and the joint-state ray view (so-named by Maudlin)), the state nonreducibility is due to the fact that the states of the subsystems do not determine the state of the composite system. (See Maudlin, pp.50-55).

To think that the dynamical behavior of a system supervenes on the description of its states and if the states are reducible, the dynamical behavior should also be reducible, is mistaken because, as I have argued in the previous paragraph, although the state of a nonseparable system can always be described by a point in phase space, which is reducible to the states of the subsystems, its motion or trajectories contain singular points that cannot be reduced to the independent trajectories of the subsystems. So the nonreducibility of the Hamiltonian and the behavior of nonseparable systems is not a mere epistemic nonreducibility, but one with ontological significance. The description of the dynamical behavior of a nonseparable system does not supervene on the description of the dynamical behavior of its subsystems.

So the dynamical holism in classical mechanics, like the state holism in quantum mechanics, is also ontological. As I have shown in the previous sections, the nonseparability of Hamiltonians and the nonreducibility of the behavior of nonseparable systems is not a matter of our limited epistemic ability to separate the Hamiltonians or to reduce the behavior of nonseparable systems. They reflect the essential dynamical properties of the systems.

(c) Objection: dynamical holism presumes the reality of "imaginary" subsystems

Another possible objection to the notion of dynamical holism is that dynamical holism concerns the nondecomposability of the motion of a system to the "motions" of *theoretical or imaginary* generalized subsystems. Since the generalized subsystems are epistemic devices, the notion of dynamical holism is not philosophically or ontologically significant.

To this possible objection, I will argue that while it is true that the generalized subsystems cited in the definition of dynamical holism are theoretical, as I will describe in detail below, this does not undermine, but rather reinforces, the ontological importance of the notion of dynamical holism.

As we have seen from the previous sections, generalized subsystems of a dynamical system can be ordinary, i.e., commonsense subsystems like interacting particles. They can also be theoretical subsystems described by generalized coordinates and momenta where each of the generalized coordinate and momentum may be some functions of the Cartesian coordinates of the particles.

Also, the generalized subsystems used in the notion of dynamical holism are degrees of freedom rather than ordinary subsystems. As we have seen, the nonseparability of a Hamiltonian is defined by the fact that it cannot be written as the sum of the Hamiltonians in each degree of freedom. It appears as the nonseparability of the pairs of the conjugate variables: each pair of the conjugate variables represents one degree of freedom. The nonreducibility of the Hamiltonian implies that the motion of the system cannot be reduced to the (independent) motions in each degree of freedom. Although the concept of generalized subsystems includes ordinary subsystems as one possible set of generalized subsystems since the ordinary subsystems can be described by one set of generalized variables, it also includes other sets of theoretical subsystems represented by sets of generalized variables resulting from canonical transformations. So the nonreducibility of the Hamiltonian and behavior of a nonseparable system means that it is not reducible to the Hamiltonian and behavior of *any* set of generalized subsystems described by a set of generalized variables (not just that it is not reducible to the Hamiltonian and behavior of its ordinary spatial subsystems described by a certain set of generalized variables).

But when it is separable, it can be reduced to at least one set of generalized subsystems. If the Hamiltonian of a system is not separable in terms of ordinary subsystems but separable in terms of some theoretical subsystems, the system is not dynamically holistic, i.e., it is reducible to the Hamiltonians of some set(s) of subsystems. This indicates that the motion of the system is regular and we can solve its equations of motion in terms of some generalized coordinates and generalized momenta. Then we can get the description of the motion of the ordinary subsystems through inverse transformation. This means that the motion of the ordinary subsystems can actually be characterized independently, that is, the $x_i(t)$'s and $p_i(t)$'s are independent functions. So whether a system is separable and therefore integrable and reducible is *not* determined by whether its Hamiltonian can be written as a sum of the Hamiltonians of its *ordinary* subsystems.

The generalized subsystems used by the notion of dynamical holism are also theoretical or imaginary in the sense that the generalized subsystems are described by generalized coordinates and generalized momenta. The generalized coordinates and generalized momenta, are functions of the positions and momenta of the ordinary subsystems. However, each pair of conjugate variables describes the motion of one degree of freedom. So each generalized subsystem represented by a pair of conjugate variables is a part of the system that contributes to the motion of the system in one degree of freedom. Therefore, the nonreducible dynamical system is not just not reducible to the dynamics of separable *ordinary* subsystems. It is a dynamical whole that is not divisible into *any* separable dynamical parts. Since each generalized subsystem represented by a pair of conjugate variables is necessary in describing the motion of the system, it can be seen as a necessary element in describing the motion of the system.

Although the generalized subsystems in the notion of dynamical holism are theoretical, it does not mean that dynamical holism is epistemic. Rather, it shows that the dynamically holistic behavior is independent of representations and therefore is ontological. If we see the different ways of dividing the system into the generalized subsystems (including the ordinary subsystems) as different representations of the system, the fact that the dynamics of the nonseparable system is not reducible to the dynamics of any of these generalized subsystems indicates that the nonreducibility of the dynamics of the system is invariant or is independent of the way the system is represented. (It is generally believed that if something is invariant or independent of representation, it has ontological significance). Conversely, if a system is not reducible to the ordinary subsystems but is nonetheless reducible, its nonreducibility is representation-dependent or epistemic and it is not really a nonseparable system.

Although the necessity of introducing these theoretical subsystems into the definition of dynamical holism can be seen from the discussions in the previous sections, to make it more explicit, I will summarize these points briefly as follows: First, the Hamiltonians of the systems are generally described in terms of generalized variables (representing the degrees of freedom), not in terms of spatial subsystems. Secondly, generalized coordinates and generalized momenta are introduced to describe effectively the state of the system and to eliminate efficiently coupled terms or interactions in order to solve the equations of motion. Without using these generalized coordinates and momenta, we would not be able to determine whether a Hamiltonian is separable or not. The system represented by the theoretical subsystems is equivalent to the system represented by ordinary subsystems or described in terms of Cartesian coordinates. But in general, the description of a system in terms of Cartesian coordinates and their conjugate ordinary momenta does not tell whether the system is separable or not. As Helleman (Helleman, p.170) notes, whether a system is integrable is determined by whether its canonical transformation series, i.e., Birkhoff series, converges or not, that is, whether it can be canonically transformed to a system whose Hamiltonian is a function of the new or transformed momenta only. The dynamical behavior of a system (such as whether its motion is regular) is determined not by whether its Hamiltonian is a sum of the Hamiltonians of the spatial subsystems, but by whether its Hamiltonian can be a sum of some generalized subsystems. Thirdly, for systems with spatial subsystems, the ordinary subsystems and theoretical subsystems are mathematically and physically equivalent in describing the motion of these systems, i.e., they are different representations of these systems. If solutions exist for ordinary subsystems, they also exist for theoretical subsystems and vice versa because in principle, we can always transform from one to the other. Similarly, if the solutions do not exist in terms of theoretical subsystems, they also do not exist in terms of ordinary subsystems. That is, if the motion is regular, it is regular in any representation and if it is chaotic, it cannot be described in terms of any canonically conjugate variables.

A final argument for introducing theoretical subsystems in the definition of dynamical holism is that the abstractness of generalized subsystems is a characteristic not only of dynamical holism in classical mechanics, but also a characteristic of dynamical holism in quantum mechanics, and more importantly, it is also contained in the concept of state or property holism. Since in quantum mechanics a nonseparable Hamiltonian is also defined as one that is not reducible to the Hamiltonians of the generalized subsystems, such generalized subsystems are used in that context as well. Moreover, as Kronz points out, a nonseparable Hamiltonian in quantum mechanics produces not only nonseparable evolution, it also generates nonseparable states or properties. So the description of the nonseparable states or properties would also involve these generalized variables.

Conclusion

From the discussion in this chapter we can conclude that dynamical holism in classical mechanics is real: the motions of nonseparable systems are holistic. An important consequence of dynamical holism demonstrated by nonseparable systems is that the dynamically holistic systems, in the context of perturbation approaches to their solutions, show a kind of nonreducibility, which is closely related to the philosophical exportation problem that I mentioned in the Introduction. Since dynamically holistic systems are not integrable, scientists use perturbation techniques to obtain approximate solutions to the equations of these systems. According to perturbation theory, although we cannot separate the subsystems of a nonseparable system, we may be able to write its nonseparable Hamiltonian as the sum of a separable one plus a small perturbation, that is, $H = H_0 + H'$ (where H_0 is a separable Hamiltonian and H' is the perturbation). If so, we hope to write the solutions for the nonseparable system as the sum of the solutions of the separable system represented by the separable Hamiltonian and small correction terms resulting from the perturbation. The solutions for the separable system will asymptotically approximate the solutions for the nonseparable system when the perturbation approaches zero. If so, we would be able to get approximate solutions for nonseparable systems from the solutions for the (solvable) separable systems. Unfortunately, perturbation theory shows that the solutions for nonseparable systems do not uniformly approximate the solutions for the unperturbed separable systems, as I will show in the next chapter. This will cause problems for the theories of exportations and idealizations, which I will discuss in detail in Chapter Three.

Chapter Two. Nickles nonreducibility of dynamically holistic systems

Introduction

As I have shown in the last chapter, in systems with nonseparable Hamiltonians, the motion of the whole system is not the sum of the motions of their generalized subsystems. Such systems are called "dynamically holistic". For these systems, we cannot find exact solutions to their equations of motion by applying the techniques of canonical transformation and integration. Rather, we have to use perturbation methods (if they are applicable) to obtain approximate solutions for them. In this chapter I will show that a consequence of dynamical holism is that when we treat dynamically holistic systems by perturbation methods, we encounter *singular* perturbation problems whose dynamical behavior cannot be approximated by that of the simplified separable unperturbed systems. This consequence has important philosophical implications because it can be seen as a failure of a kind of reduction which is closely related to the theories of exportation and idealization.

The connection between perturbation theory and theories of exportation and idealization may be seen from the fact that perturbation methods can be taken as ways of describing how idealized or simplified systems are related to real systems. Perturbation theory describes real systems as perturbed systems and idealized systems as unperturbed systems. Based on perturbation theory, Nickles developed a limit notion of reduction. It says that if a theory of a complicated system converges to a theory of a simplified system in the limit, which in terms of perturbation theory means that the complicated system is a regular perturbation of the simplified system, the former can be reduced to the latter. Nickles reducibility provides an apparent basis for justifying the scientific practice of exporting idealized theories (or descriptions of simplified models) to real cases: since the description of a real system can be reduced to the description of an idealized system, idealized theories describing idealized systems can be used to describe, at least approximately, real systems. As I will discuss in detail in the next chapter, some accounts of scientific exportation explicitly or implicitly assume that for the descriptions of idealized systems to be exportable to real systems, the real systems must be Nickles reducible to the idealized systems. But if dynamically holistic systems involve singular perturbation problems and Nickles reduction fails in such cases, the accounts of exportation based on Nickles reduction would not succeed.

Below I will first introduce a more formal definition of Nickles reduction. The ensuing discussion will require the introduction of some basics of perturbation theory and the important distinction between regular and singular perturbations. Finally, I will show that dynamically holistic systems are singular perturbation problems which are not Nickles reducible to their unperturbed systems. Since these unperturbed systems can be seen as the "main" subsystems of dynamically holistic systems, dynamically holistic systems are not Nickles reducible to their main subsystems.

2.1. The notion of Nickles reduction

Nickles (Nickles, 1973, p.182) defined a notion of reduction to capture the general practice of using perturbation theories in physics and mathematics. It is different from the popular notion of reduction (due to Nagel) that defines a theory as reducible to another one if it is derivable from the latter. Nickles defines a complex theory as reducible to a less complex one if the complex theory becomes the simpler theory in the limit as the small parameter in the complex theory goes to zero or some special value. Usually, the complex theory can be seen as describing a perturbed system, employing a small parameter to represent the strength of the perturbation. Formally, Nickles' notion of reduction can be expressed as follows: a theory Θ_2 (which is a function of a variable x and a parameter ε) is Nickles reducible to Θ_1 if

$$\lim_{\varepsilon \to 0} \Theta_2(x, \varepsilon) = \Theta_1(x) \qquad \text{for all } x$$

which means that Θ_2 should converge to Θ_1 for all values of x in the domain where the theories are defined, as the parameter(s) ε approaches *0* (or some other suitable value). We may express Nickles reduction in terms of perturbation theory: if we see Θ_2 as the description of a perturbed dynamical system and Θ_1 as the description of the unperturbed system, we can say a system is Nickles reducible to its unperturbed system if its motion converges to that of the unperturbed system for all values of x as the perturbation approaches zero.

As an example of Nickles reduction, we note that some parts of special relativity theory can be Nickles reduced to Newtonian mechanics in the limit of low velocities, $\frac{v^2}{c^2} \rightarrow 0$. For example, $\lim_{\frac{v^2}{c^2} \rightarrow 0} m_0 v / \sqrt{1 - \frac{v^2}{c^2}} = m_0 v$. That is, the relativisitic momentum is reduced to the classical momentum in the limit of low velocity (see Batterman, 1995).

As its definition indicates, Nickles reduction reduces a "finer" or general theory to a "coarser" or less general theory since the parameter in the reduced theory can have various values whereas it can only have a specific value such as 0 in the reducing theory. From a methodological point of view, Nickles reduction supports the reduction of a dynamical description of a complex system to the dynamical description of a less complex system when the limiting condition is met. Nickles' notion of reduction is apparently defined as reduction of theories, which normally would be identified with the reduction of the *equations* of theories. However, as Batterman argued, for a theory to be reducible to another theory not only must the *equations* of the latter be the limit of the *equations* of the former for all the values of the variable, but the *solutions* of the latter must be the limits of the *solutions* of the former for all the values of the variable (Batterman, 2003). This is because if the solutions are not Nickles reducible, even if the equations of a system can become the equations of a simpler system in the limit of a special value of the parameter(s), the reducing and reduced equations describe systems with very different behaviors. As I will show, when the solution of a system is not Nickles reducible to the solution of its main subsystem, not only does the latter not approximate the former quantitatively, it even has qualitatively different behavior than the former. Therefore, if the equations of a theory can be Nickles reduced to those of another theory but the solutions to the equations cannot be Nickles reduced to the solutions to the equations of the other theory, the former theory is not Nickles reducible to the latter.

If we apply the notion of Nickles reduction to perturbation problems, we will find that the so called regular perturbation problems are Nickles reducible whereas the singular perturbation problems are not reducible to their unperturbed problems. But to show how perturbation theory is related to Nickles reduction (in order to illustrate that dynamically holistic systems when treated as perturbation problems are not Nickles reducible to their unperturbed or idealized systems), I need first to introduce some basics of perturbation theory and the important distinction between regular and singular perturbation problems.

2.2. Some basic concepts of perturbation theory

The central idea of perturbation theory is to get approximate solutions to dynamical problems by solving the unperturbed problem exactly and adding small corrections terms to the solution. The construction of the approximate solution of perturbation problems is based on the fact that the perturbation problems can be simplified or idealized as unperturbed problems which have exact solutions. The difference between the perturbed and the unperturbed problems is the perturbations which are expressed in the perturbation equations as terms containing a small parameter. For example, the van der Pol equation

$$\ddot{x} - \varepsilon(1 - x^2)\dot{x} + x = 0$$

can be simplified as the unperturbed equation

$$\ddot{x} + x = 0$$

The difference between the perturbed equation and the unperturbed equation is the small perturbation term $\varepsilon(1 - x^2)\dot{x}$. Some nonseparable systems such as the system of the planets, can be described by a set of perturbation equations such as the following (in action-angle variables)(Arnold, 1983, p.139):

$$\dot{I}_j = \varepsilon f_j(I, \varphi, \varepsilon)$$

$$\dot{\varphi}_j = \omega_j(I) + \varepsilon g_j(I, \varphi, \varepsilon) \qquad (j = 1, 2, \ldots n)$$

where the terms $\varepsilon f_j(I, \varphi, \varepsilon)$ and $\varepsilon g_j(I, \varphi, \varepsilon)$ describe small perturbations. These perturbation equations can be idealized (by reducing the perturbations to zero) as the equations

$$\dot{I}_j = 0$$

$$\dot{\varphi}_j = \omega_j(I) \qquad (j = 1, 2, \ldots n)$$

The idealized equations are integrable. They describe a separable system. The small parameter ε measures the magnitude or strength of the perturbation. As we will see, perturbation theory requires the perturbations to be small, that is, the perturbation parameter must be small in order for the solutions of perturbation theory to approximate the true description of perturbed systems. The perturbation parameter indicates how close the real perturbed (nonseparable) system is to the unperturbed ideal (separable) system. This provides the basis for solving the perturbed problem: to construct the solution for the perturbed system based on the solution for the unperturbed system, which may be regarded as the "*main subsystem*" of the perturbed system.

Perturbation theory constructs the solution of the perturbed problem by an asymptotic expansion, which includes the solution of the unperturbed problem and some modifications or corrections of it in terms of the perturbation parameter. The asymptotic expansion allows us to take the sum of the first few terms (or even the first term) in the expansion series as an approximate solution of the perturbed problem. This means that if the solution of the unperturbed problem is the first term of the expansion, it can be seen as an approximate solution to the perturbed problem.

We must note that in order for the first few terms to be an approximate solution, the expansion must be *asymptotic* rather than *convergent.* We might mistakenly think that in order for an expansion to be an approximate representation of the solution of a perturbed system, it should converge to the real solution. This means that when we add more and more terms in the expansion we will get a more and more accurate result and when the number of terms goes to infinity, we get the exact solution. So we may expect the approximate solution to be a convergent expansion which is defined as the following (see Hinch, p.19): an expansion $\sum_{n=1}^{\infty} x_n(t)$ is convergent to $x(t)$, that is,

$$\lim_{N\to\infty} \sum_{n=1}^{N} x_n(t) = x(t)$$

at a fixed value of *t*, if (and only if) given an arbitrary $\delta>0$, it is possible to find a number $N_0(t,\delta)$ such that

$$\left|[\sum_{n=1}^{N} x_n(t)] - x(t)\right| \le \delta$$

for all $n \ge N_0$

If we look for convergent approximate solutions for perturbed systems, we would need to find convergent expansions that can represent the solutions. Such expansions would be parameterized by the small perturbation parameter ε, that is, they would be described as $\sum_{n=1}^{N} x_n(t,\varepsilon)$ or $\sum_{n=1}^{N} (x_\varepsilon)_n(t)$. Such expansions could be convergent in one of the following two ways: If the expansion converges to the limiting function $x_{\varepsilon=0}(t)$ as $\varepsilon \to 0$, $N \to \infty$ for each value of *t*, the expansion is called *pointwise convergent*. This is because the $N_0(t,\delta)$ in such convergence could be different for each value of *t*. In contrast if there is a $N_0(\delta)$, independent of *t*, such that the expansion converges to the limiting function $x_{\varepsilon=0}(t)$ for *all* the values of *t* in its domain as $\varepsilon \to 0$, $N \to \infty$, we say the expansion is *uniformly convergent*. As we will see, the distinction between pointwise and uniform convergence is crucial for the distinction between regular and singular perturbation, which is in turn very important for the notion of Nickles reducibility/nonreducibility of dynamical systems.

The definition of convergence above implies that if an expansion as an approximate representation of *x*(t) is convergent, we can make it to any desired accuracy by truncating the expansion at some point. However, convergent expansions may not be useful in the practice of getting approximate solutions because we may need many terms in the expansion to approximate the true description. For example, the solution of Bessel's equation of order zero, *x*(*t*), can be represented by two expansions (Nayfeh, pp. 23-24 and pp. 340-342). The first one is

$$x(t) = 1 - \frac{t^2}{2^2} + \frac{t^4}{2^2 \cdot 4^2} - \frac{t^6}{2^2 \cdot 4^2 \cdot 6^2} + \cdots$$

The second one is

$$x(t) \sim \sqrt{\frac{2}{\pi t}}\left[u(t)\cos\left(t - \frac{1}{4}\pi\right) + v(t)\sin\left(t - \frac{1}{4}\pi\right)\right] \qquad \text{as} \quad t \to \infty$$

(The symbol "~" means that the expansion on the right hand side is an asymptotic representation of $x(t)$.)

where

$$u(t) = 1 - \frac{1^2 \cdot 3^2}{4^2 \cdot 2^2 \cdot 2!t^2} + \frac{1^2 \cdot 3^2 \cdot 5^2 \cdot 7^2}{4^4 \cdot 2^4 \cdot 4!t^4} + \cdots$$

$$v(t) = \frac{1}{4 \cdot 2t} - \frac{1^2 \cdot 3^2 \cdot 5^2}{4^3 \cdot 2^3 \cdot 3!t^3} + \cdots$$

The first expansion $x(t)$ is uniformly convergent for all values of t, (but as we will see it is not asymptotic), whereas the second expansion is divergent for all values of t. For small values of t, the first nine terms of the first expansion gives a result correct to 11 significant figures. But as t increases, the number of terms needed to get the same accuracy increases rapidly. In contrast, the first term of the second expansion provides an accuracy the same as eight terms of the convergent expansion when t=4. For very large values of t, the convergent expansion is useless from the computational point of view. But the second expansion can still give quite accurate results with just the first or the first few terms (see Hinch pp.19-20 for another example). This property of the second expansion is asymptoticness, which is what we need in practice.

The property of asymptoticness can be defined as the following (see Hinch, 1991, p.20): an expansion $\sum_{n=1}^{N} x_n(t;\varepsilon)$ is an asymptotic approximation to $x(t,\varepsilon)$ as $\varepsilon \to 0$ if for each $M \le N$,

$$\frac{x(t,\varepsilon)-\sum_{n=1}^{M} x_n(t;\varepsilon)}{x_M(t,\varepsilon)} \to 0 \qquad \text{as } \varepsilon \to 0 \ \text{ for all } t$$

This means that asymptoticness requires that the series approximate $x(t,\varepsilon)$ with a fixed number of terms. As Nayfeh puts it: if the series is to be useful, the number of terms must be fixed (Nayfeh, 1981, p.20). So the requirement for being an asymptotic series is that for any fixed number M of terms, where $M \leq N$, (N fixed), that are left after truncating the expansion, the sum of the remaining truncated terms is smaller than the last term left in the expansion as $\varepsilon \to 0$. That is,

$$\sum_{M+1}^{N} x_n(t;\varepsilon) = o[x_M(t;\varepsilon)] \qquad \text{as} \quad \varepsilon \to 0$$

(The symbol "*o*" is used in the above equation roughly to indicate "the left hand side tends to zero faster than the right hand side". [12])

So the error committed in truncating the series of asymptotic expansions is by definition the order of the first neglected term and, hence, it tends rapidly to 0 as

---

[12] The symbol "*o*" (the small *o*) means that if $\lim_{\varepsilon \to 0} \frac{f(\varepsilon)}{g(\varepsilon)} = 0$, $f(\varepsilon) = o[g(\varepsilon)]$ as $\varepsilon \to 0$. It indicates that $f(\varepsilon)$ tends to zero faster than $g(\varepsilon)$ as $\varepsilon \to 0$. In contrast, the symbol "O" (the big O, which will be used below), means "the order of", and is defined as: if a function $f(\varepsilon)$ is the order of $O[g(\varepsilon)]$, $f(\varepsilon)$ tends to zero at the same rate that $g(\varepsilon)$ tends to zero. That is, $\lim_{\varepsilon \to 0} \frac{f(\varepsilon)}{g(\varepsilon)} = a$, $0<|a|<\infty$. So $\sin\varepsilon = O(\varepsilon)$ which means that $\sin\varepsilon$ tends to zero at the same rate that ε tends to zero (Nayfeh 1981, p.17). So for an asymptotic expansion, the first term is of the order of $O(1)$, the second term is of the order of $O(\varepsilon)$ and the later terms are higher orders. So the first term can be taken as an approximate description of the expanded function as $\varepsilon \to 0$.

$\varepsilon \to 0$ (ibid., pp. 21and 23). Since M can be any number including 1, this requires that the decay is not only fast but also starts from the second term. In contrast, the convergent expansion only requires that the terms eventually decay sufficiently rapidly, as it is defined as the expansion that given an arbitrarily small number δ, can be truncated at some point (such as at the term N_0) such that the truncated part is smaller than δ. So for an asymptotic expansion, if we truncate the series after the first term, which is called the leading term $x_0(\varepsilon)$, the error we commit $o[x_0(\varepsilon)]$ is and it tends to 0 as $\varepsilon \to 0$. Thus we can say that the "leading term of an asymptotic expansion is roughly correct and further terms are small corrections of decreasing size" (Hinch, 1991, p.20). In practice, we "rarely determine more than one or two terms in the expansion" (Nayfeh, 1981, p.22). Asymptotic expansion requires that for fixed N, when the perturbation parameter $\varepsilon \to 0$, the sum of the first N (≥ 1) terms represents an asymptotic approximate solution. That is, $x(t) \sim \sum_{n=1}^{N} x_n(t;\varepsilon)$ as $\varepsilon \to 0$ for any N (≥ 1). In contrast, the convergent expansion requires that for a fixed t as $N \to \infty$, the expansion will approximate the real description. Thus, a convergent expansion may require many terms to get close to the real description.

As I have shown above, the notion of asymptoticness is very different from the notion of convergence. An asymptotic expansion can be divergent and a convergent expansion may be non-asymptotic as we have seen from the example of the two expansions of Bessel's function. In practice, what we need are asymptotic approximations to the real descriptions, although typically we need just the first term (or first two terms) of an asymptotic expansion to be the approximation. However, there is a way of connecting convergence with asymptotic approximation: when the expansion $x(\varepsilon)$ is written as a power series of the small parameter ε, that is, $x(t;\varepsilon)$, for example, $x(t;\varepsilon) = x_0(t) + \varepsilon x_1(t) + \varepsilon^2 x_2(t) + \cdots \varepsilon^n x_n(t)$, if the expansion uniformly converges to $x_0(t)$, that is, $\lim_{\varepsilon \to 0} x(t;\varepsilon) = x_0(t)$ for all t, which implies that $\sum_{n=1}^{N} \varepsilon^n x_n(t) = o[x_0(t)]$ (note that the sum does not include the first term, i.e., the $n = 0$ term) as $\varepsilon \to 0$, i.e., at least the first term of the expansion $x(\varepsilon)$ is an asymptotic approximation. Conversely, if the expansion $x(t,\varepsilon)$ does not converge uniformly to $x_0(t)$, then it is possible to find a $\delta_0 > 0$ so that no matter how small ε is, there exists a $t = t(\varepsilon)$ and the error $|x(t(\varepsilon),\varepsilon) - x_0(t(\varepsilon))| \geq \delta_0$. This means that the first term of the expansion (which is $x_0(t)$) is not a uniform asymptotic approximation of $x(t)$ and the expansion $x(t,\varepsilon)$ is not uniformly asymptotic. When the expansion converges uniformly to its first term as $\varepsilon \to 0$, the first term is called its asymptotic limit. If the first term $x_0(t)$ is the solution of the unperturbed problem, and if the expansion of the solution of the perturbed problem converges uniformly to $x_0(t)$, i.e., $\lim_{\varepsilon \to 0} x(t;\varepsilon) = x_0(t)$, the solution of the unperturbed problem can be seen as the approximation of the solution of the perturbed problem. But if the convergence fails, the solution of the unperturbed problem does not approximate that of the perturbed problem.

The limit relation between the solution of the unperturbed problem and that of the perturbed problem is the same as the limit relation between the reducing and the reduced theories in the notion of Nickles reduction. So we can say when the solution of the perturbed problem is the asymptotic limit of the solution of the perturbed problem, the solution of the perturbed problem (which describes the dynamical behavior of the perturbed system) is Nickles reducible to that of the unperturbed problem. But when the limit relation fails, the dynamical behavior of the perturbed system is not Nickles reducible to that of the unperturbed system. So when the solution of a perturbed problem is Nickles reducible to that of the unperturbed problem, the latter is a valid approximation of the former. Conversely, when the former is not Nickles reducible to the latter, the latter does not approximate the former.

Unfortunately, the asymptotic solutions for some perturbed problems, such as nonseparable problems, do not have the solutions of their unperturbed problems as their asymptotic limits. That is, the solution of the perturbed problem does not converge uniformly to the solution of the unperturbed problem. When we try to write asymptotic expansions for the solutions of these problems with the solutions of the unperturbed problems as the leading terms, the asymptotic expansions break down. This means that the solutions of the perturbed problems do not converge to and cannot be approximated by the solutions of the unperturbed problems. These problems are called singular perturbation problems. Below I will show in technical terms the distinction between regular and singular perturbation problems and explain why singular perturbation problems are not Nickles reducible to their unperturbed problems.

2.3. The distinction between regular and singular perturbation problems and the Nickles nonreducibility of singular perturbation problems

To find asymptotic solutions for perturbed systems such as nonseparable systems (which are treated as perturbed problems in perturbation theory), the simplest way is to write the solutions of them as power expansions based on the solutions of the unperturbed systems. That is, we write the asymptotic solution of a perturbed system in the following simplest form

$$x(t;\varepsilon) = x_0(t) + \varepsilon x_1(t) + \varepsilon^2 x_2(t) + \cdots + \varepsilon^n x_n(t) \quad n = 1,2\cdots \tag{1}$$

This expansion is called a regular or straightforward expansion. We can see that the leading term $x_0(t)$ is not a function of the perturbation parameter ε; thus it contains no perturbation effect. It is the solution of the unperturbed problem $x_{\varepsilon=0}(t)$. So the regular expansion of the solution of a perturbed problem consists of the solution of the unperturbed problem as its leading term plus small corrections of the order ε and higher orders. If as $\varepsilon \to 0$, the solution of the perturbed problem, $x(t;\varepsilon)$ uniformly converges to or becomes $x_0(t)$, the solution of the unperturbed problem, the solution of the unperturbed problem is an asymptotic limit of the perturbed problem, that is,

$$\lim_{\varepsilon \to 0} x(t;\varepsilon) = x_0(t) \qquad \text{for all values of } t \tag{2}$$

in the relevant domain of t (O'Malley, pp.4-5).

Here the requirement of "for all values of t" is crucial for the distinction between a uniform and nonuniform (or a valid and nonvalid) asymptotic approximation. For an expansion $x(t;\varepsilon)$ to be a valid approximation of $x(t)$, it has to be a uniformly asymptotic series, that is, it has to be asymptotic (with respect of ε) for all values of t. When it fails to be so, it is called a nonuniform asymptotic approximation and it does not converge to its leading term of a regular expansion uniformly, i.e., for all values of t, (given ε) but only pointwise.

Therefore, the solutions of the unperturbed problem are not the valid asymptotic approximate solutions of the perturbed problem because the solutions of the unperturbed problem are singular rather than uniform (asymptotic) limits of the solutions of the perturbed problem as $\varepsilon \to 0$. This means that as $\varepsilon \to 0$, the solutions of the perturbed $x(t,\varepsilon)$ converge to $x_{\varepsilon=0}(t)$ only pointwise, i.e., only for some values of t. For some t, even if ε is arbitrarily close to zero, $x(t;\varepsilon)$ is radically different from $x_{\varepsilon=0}(t)$. If we expand $x(t,\varepsilon)$, which has the solution of the unperturbed problem as a singular limit, in the form of a regular expansion, we will get a non-uniform (or nonasymptotic) expansion. That is, the asymptoticness of the expansion will break down for some values of t. But an asymptotic expansion has to be uniformly asymptotic for the values of t that we are concerned with to be an approximate description of the perturbed system.

The perturbed problems that have the unperturbed solutions as singular limits or have non-uniform regular expansions, are called singular perturbation problems. On the other hand, if the asymptoticness of the regular expansion does not break down and thus the solution of the unperturbed problem is the uniform limit of the solution of the perturbed problem, the perturbed problem is a regular perturbation problem. We may follow O'Malley (O'Malley, pp.4-5) to define the notion of a regular perturbation problem as: if the solution of the perturbed problem, say $x(t,\varepsilon)$ uniformly converges to the solution of its limiting or unperturbed problem, $x_{\varepsilon=0}(t)$ as the small perturbation parameter $\varepsilon \to 0$, that is,

$$\lim_{\varepsilon \to 0} x(t,\varepsilon) = x_{\varepsilon=0}(t) \qquad \text{for all } t$$

the perturbed problem is a regular perturbation problem, and if the above regular perturbation limit fails, the perturbed problem is a singular perturbation problem.

Since as $\varepsilon \to 0$, the *equations* of motion of a perturbed system always (in all cases) become the equations of the unperturbed system, this may lead us to think that the *solutions* of the equations of the perturbed system can also always approximate the solution of the equations of the unperturbed system. We may then expect that the leading term of the perturbed system is always the solution of the unperturbed system $x_{\varepsilon=0}(t)$ so that the solution of the perturbed system can always be approximated by the solution of the unperturbed system in the limit $\varepsilon \to 0$. This idea provides the basis for the generally received view of scientific methods, which takes the descriptions of the real systems (perturbed problems) to be approximated by the descriptions of the ideal systems (unperturbed problems) as I will show in detail in the next chapter. However, for the general perturbation problems such as those of the nonseparable systems, although the equations of the perturbed systems can be approximated by the equations of the unperturbed systems in the limit, the solutions of the equations of the perturbed systems cannot be approximated by the solutions of the equations of the unperturbed systems. In other words, although the unperturbed problem is the limiting problem of the perturbed problem, the solutions of the unperturbed problem are not the approximate solutions of the perturbed problem.

This is because for most dynamical problems, the regular perturbation procedure fails to get uniform valid asymptotic expansions. There can be many reasons for this failure. One of them is that ε multiplies the highest derivatives in the differential equation. For example, the heavily damped van der Pol oscillator

$$\varepsilon\ddot{x} - (1 - x^2)\dot{x} + x = 0$$

is a case of this type. The limiting or unperturbed problem for this *second order* differential equation is

$$x - (1 - x^2)\dot{x} = 0$$

which is a *first order* differential equation. Another reason for regular perturbation theory to fail is that the governing equation contains small terms with a cumulative effect. For example, the lightly damped van der Pol equation mentioned above

$$\ddot{x} - \varepsilon(1 - x^2)\dot{x} + x = 0$$

In the limit, as $\varepsilon \to 0$, the solution of the limiting case (i.e., the unperturbed problem) is not the uniform limit of the solution of the perturbed problem. The breakdown of the asymptoticness, as we will see in the example below[13], is due to the terms in the expansion which become very large when t is large, even when ε is very small. These terms are called *secular* terms implying that after very long times they become non-negligible.

Consider an oscillator with a nearly constant frequency ($\sqrt{1-\varepsilon e^{-at}}$) that can be described by the following equation

$$\ddot{x} + (1 - \varepsilon e^{-at})x = 0 \tag{3}$$

where a is a positive constant, with the initial conditions:

$$x(0) = 0$$
$$\dot{x}(0) = 1$$

The regular perturbation procedure is to substitute the regular expansion (1) into the differential equation. By collecting the terms with the same order of ε on both sides of the equation; we get a series of equations, which can be recursively solved and determines the terms $x_0(t)$, $x_1(t)$, $x_2(t)$ etc. If the solution is uniform, this problem is a regular problem and the solution is a valid approximation to the true solution. The accuracy of the result improves as ε gets smaller.

---

[13] The example is from Kevorkian and Cole (1996) p.21ff.

So by substituting the regular expansion (1) into the equation and comparing the coefficients of ε^n, we get a sequence of problems.

At ε^0, $\ddot{x}_0(t) + x_0(t) = 0$ (4)

The above equation (for the leading term $x_0(t)$ in the regular expansion) is the same as the equation of the limiting case, i.e., the unperturbed problem.

The solution of (4) (for the given initial conditions) is

$$x_0(t) = \sin t$$

At ε^1, we have, $\ddot{x}_1(t) + x_1(t) = e^{-at}\sin t$

The solution of the above equation is

$$x_1(t) = \frac{e^{-at}+1}{4+a^2}\sin t + 2\frac{e^{-at}-1}{a(4+a^2)}\cos t + O(\varepsilon^2)$$

Thus,

$$x(t,\varepsilon) = \sin t + \varepsilon[\frac{e^{-at}+1}{4+a^2}\sin t + \frac{2(e^{-at}-1)}{a(4+a^2)}\cos t] + O(\varepsilon^2)$$

This expansion is uniformly valid. However, if the term $-\varepsilon e^{-at}$ is replaced by εx^2, for example, which does not decay as $t \to \infty$, the equation (3) becomes

$$\ddot{x} + (1-\varepsilon x^2)x = 0$$

which is the so called Duffing equation describing a weakly nonlinear oscillator without damping. With the same initial conditions, if we apply the regular perturbation procedure to the Duffing equation, we get

$$x_0(t) = \sin t$$

but

$$x_1(t) = -\frac{9}{32}\sin t - \frac{1}{32}\sin 3t + \frac{3}{8}t\cos t$$

Thus,

$$x(t,\varepsilon) = \sin t - \varepsilon(\frac{9}{32}\sin t - \frac{1}{32}\sin 3t + \frac{3}{8}t\cos t) + O(\varepsilon^2)$$

The above expansion is not uniformly asymptotic because of the correction term $\frac{3}{8}\varepsilon t\cos t$, which becomes of the order of the leading term when t is the order of ε^{-1}. This term is a secular term which breaks down the asymptoticness of the expansion. Thus, the solution of the Duffing equation does not uniformly converge to its regular leading term $x_0(t)$, which is the solution of its limiting case $x_{\varepsilon=0}(t)$. The problem is therefore, a singular perturbation problem.

In order to get uniform asymptotic expansions for singular problems, as we will see in Chapter Four, we have to use multiple scales expansions to replace the one-scale regular expansions. Consequently, the solutions of the unperturbed problems will not be the leading terms (and therefore, not the uniform limits) of the asymptotic expansions for the singular perturbed problems

So the fact that for singular perturbation problems the solutions do not go over to the solutions of the unperturbed problem in the limit can be seen as a failure of Nickles reduction. So all singular perturbation problems are not Nickles reducible to their unperturbed problems. Conversely, all perturbed problems that are Nickles reducible to the unperturbed problems, are regular perturbation problems.

If we take the unperturbed system as a major subsystem of the perturbed system (since it is contained in the latter and is simpler than the latter), the non-reducibility of singular perturbation problems means the description of the perturbed problem cannot be Nickles reduced to the description of its main subsystem. So the nonreducibility of singular perturbation problems implies that we cannot approximate the description of singular perturbation problems by the description of their main subsystems.

The Nickles nonreducibility of singular perturbed problems to their main subsystems raises problems for the theories of exportation and idealization because, as I will show below, dynamically holistic systems, when treated by perturbation methods, have to be singular perturbation problems and thus are not Nickles reducible to their unperturbed systems or main subsystems. According to the received view of exportation and idealization, exportation is justified when the idealized description is the limit of the descriptions of real systems, i.e., real systems are Nickles reducible to idealized systems. Thus, the failure of Nickles reducibility of dynamically holistic systems means that we cannot export our descriptions of unperturbed systems to dynamically holistic systems.

2.4. Nonseparable systems can be treated only by singular perturbation theory

We may see that nonseparable systems have to be singular perturbation problems for the following reason: nonseparable systems have divergent perturbation (or Birkhoff) series. When we try to transform the description of a nonseparable system of n degrees of freedom to the description in terms of new variables such as action-angle variables, we write each of the old variables $q_1,\cdots q_n, p_1,\cdots p_n$ as a power series in all the new variables, say $I_1,\cdots I_n,\theta_1,\cdots\theta_n$:

$$q_i = Q_i + \varepsilon T_i(Q_1,\cdots Q_n, P_1,\cdots P_n) + \varepsilon^2 U_i(Q_1,\cdots Q_n, P_1,\cdots P_n) + \cdots$$

$$p_i = P_i + \varepsilon L_i(Q_1,\cdots Q_n, P_1,\cdots P_n) + \varepsilon^2 M_i(Q_1,\cdots Q_n, P_1,\cdots P_n) + \cdots$$

These power series are called Birkhoff series. If the Birkhoff series of the system all converge, the system is integrable. If the system is nonintegrable (i.e., nonseparable), it must have divergent Birkhoff series (see Helleman, 1980, p.170).

However, according to the definition of regular perturbation, the solution of a regular perturbation problem must converge to the solution of the unperturbed problem. This means that a regular perturbation problem must have convergent regular perturbation series. If we see regular perturbation series as a special case of Birkhoff series we can say that a nonseparable system, if it can be treated as a perturbation problem, cannot be a regular perturbation problem but has to be a singular perturbation problem. This means that the solution or behavior of a nonseparable system cannot be Nickles reduced to the solution or behavior of the unperturbed system (or its main subsystem).

We should note that although all nonseparable systems when treated as perturbation problems are singular perturbation problems, not all singular perturbation problems are nonseparable problems. Some singular perturbation problems are integrable, when the perturbation expansions of the singular perturbation problems are convergent. The systems associated with these problems are dynamically atomistic. This indicates that if a system is dynamically holistic, which means that it is nonintegrable, it entails that it is not Nickles reducible to its unperturbed system (since nonintegrable problems are singular problems), but the reverse does not hold. When the perturbation expansion of a singular perturbation problem is convergent, although the expansion does not uniformly converge to the solutions of the unperturbed problems (since they are singular problems), the equations of motion of the system are integrable. This means that although the perturbation expansion of the system are not Nickles reducible to the solutions of its unperturbed system in the limit, the solutions of the equations of motion of the system can be the "sum" of the solutions of some (canonically transformed) separable generalized subsystems. So some problems that are not Nickles reducible to the unperturbed problems are in principle solvable. Conversely, all Nickles reducible problems are regular perturbation problems and thus are integrable. But not all separable problems are Nickles reducible to the corresponding unperturbed problems since some integrable problems can be singular perturbation problems.

We may use a diagram to illustrate the relation between separability and Nickles reducibility:

Nonseparability ⇔ Nonintegrability ⇔ Dynamical holism → Singular perturbation problem → Nickles nonreducibility

The symbol "→" means "entails".

An important consequence or implication of Nickles nonreducibility is that not only can the solutions of the dynamically holistic problems not be approximated by the solutions of the unperturbed problems *quantitatively*, the solutions are *qualitatively* different as well. This raises problems for the accounts of exportation based on the qualitative similarity between the real systems and the idealized models.

2.5 Qualitative irreducibility of Nickles nonreducible systems

Some may argue that the scientific practice of exportation can withstand the failure of nonseparable systems to be Nickles reducible. It might be possible for the solutions of the unperturbed equations to correctly describe the *qualitative* behavior of the nonseparable systems even if they differ *quantitatively*. Some philosophers have argued that scientific theories describe the qualitative properties or behavior of the systems. For example, Rueger and Sharp (1998) argue that what scientific methods require for a scientific theory to successfully describe the systems in the real world is that the qualitative behavior of the systems described by the theory be stable against perturbations.[14] Even if a theory could approximately describe the behavior of a real system quantitatively, if it fails to describe the qualitative behavior of the system correctly, it is not the right description.[15] For example, the trajectories of the damped oscillator can be quantitatively very close to those of the undamped oscillator when the damping is small enough, but since their trajectories are topologically different and thus they have qualitatively different behaviors, the behavior of the damped oscillator cannot be reduced to that of the undamped one.

If the Nickles nonreducibility of singular perturbation problems is merely quantitative but not qualitative, it might not be very important for the description of the dynamical behavior of the systems according to their viewpoint. However, although the Nickles nonreducibility is expressed as the quantitative irreducibility of the solutions of perturbed problems to the unperturbed ones, it is not merely a quantitative nonreducibility. Rather, it is closely related to qualitative nonreducibility.

---

[14] As Rueger and Sharp (1998) note, Zeeman (1988), Shirer (1987) and Guckenheimer and Holmes (1983) all argue for structural stability to be a requirement for a good theory.

[15] Liu (1999) argues a similar point, claiming that scientific theories or laws aim for lawlikeness (qualitative closeness) not truthlikeness (quantitative closeness).

As we saw in Section 1.4 (a), the qualitative nature of the dynamical behavior of a system is defined by the topology of its phase portrait. Thus, the behavior of a singularly perturbed system is qualitatively different from that of its unperturbed system if the systems have topologically different phase portraits. In the following, I will use an example to show that singular perturbation problems do have phase portraits different from those of their unperturbed problems.

Let's still use the Duffing equation, which is a singular perturbation problem. The phase portrait of the system described by the Duffing equation is: when $\varepsilon>0$, its trajectories are concentric ellipses, which correspond to periodic motion, but when $\varepsilon<0$, its phase portrait consists of ellipses, separatrices and trajectories opening to the right and trajectories opening to the left. In contrast, the phase portrait of its unperturbed system consists only of concentric ellipses (see Figure 3). Obviously, the topology of the unperturbed system is not equivalent to that of the perturbed system, that is, no smooth deformation can transform the set of trajectories of the unperturbed system into those of the perturbed system.[16] The two topologies are significantly different, no matter how small ε is. This shows that the solution of the

---

[16] It seems trivial that the topologies of the Duffing problem and its unperturbed problem are different since one has two different phase portraits in different domains whereas the other has only one phase portrait. But in other examples such the van der Pol problem, even if the singularly

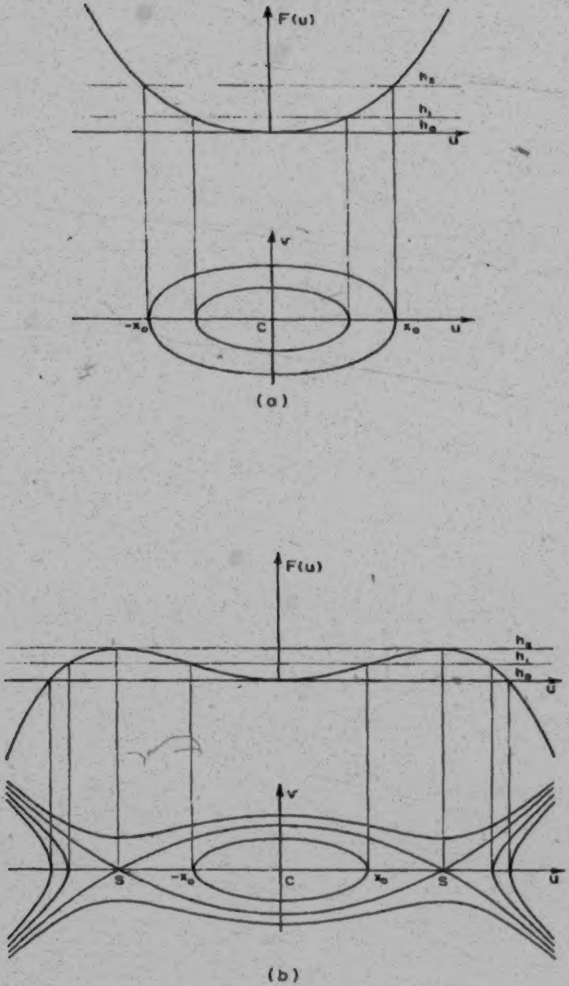
F(u)
h₂
h₁
h₀
u
v
-x₀
C
x₀
u
(a)
F(u)
h₂
h₁
h₀
u
v
S
-x₀
C
x₀
S
u
(b)

Figure 4-1. Phase plane for the Duffing equation: (*a*) $\epsilon > 0$; (*b*) $\epsilon < 0$.

Figure 3

From A. H. Nayfeh, 1981, P.115, figure 4-1

---

perturbed and the unperturbed problems both have only one phase portrait, one cannot be transformed into the other.

singular perturbation problem cannot be reduced to the solution of its unperturbed problem qualitatively in the limit.

Although it seems that the Nickles nonreducibility of singular perturbation problems to the corresponding unperturbed problems is closely related to the topological or qualitative difference between their behaviors, it seems that this has not been mathematically proved. However, it is certain that the singular perturbation problems resulting from treating nonseparable systems as perturbation problems, are qualitatively irreducible to their unperturbed problems. Nonseparable systems must be treated as singular perturbation problems, but they are a special kind of singular perturbation problem: they have chaotic motion, which, as noted in Section 1.4 (a), implies their topological or qualitative difference from the separable (or the unperturbed) systems. So the solutions of nonseparable systems are not only quantitatively but also qualitatively irreducible to the solutions of the unperturbed systems and the qualitative difference between the nonseparable systems and their unperturbed systems in the long run can lead to significant quantitative difference between them.

We may note that the qualitative difference between the behaviors of Nickles nonreducible systems and their main subsystems is not exactly the same as the qualitative difference between the behavior of dynamically holistic systems and that of the systems composed of the "artificial" independent subsystems (described in Section 1.4 (a)). Although both of the qualitative differences are topological differences, we should note the two senses of qualitative irreducibility have different meanings. Whereas Nickles qualitative irreducibility is related to the singularity of the behavior of dynamical systems at some special value(s) of the perturbation parameter (very small changes of the parameter around this special value would lead to qualitatively different behavior), dynamical holism in qualitative behavior is related to the nondecomposability of the trajectories of the nonseparable systems. The qualitative difference between dynamically holistic systems and the separable systems can be seen as the difference between chaotic motion and regular motion whereas the qualitative difference between the Nickles nonreducible systems and the unperturbed systems can be seen as the difference between two topologically nonequivalent regular motions.

The relation between the two senses of qualitative irreducibility is like the relation between dynamical holism and Nickles nonreducibility: nonseparable systems that behave qualitatively differently from the sum of their "artificial" independent subsystems are also qualitatively different from their unperturbed systems. That is, the behavior of dynamically holistic systems is qualitatively different from that of the sum of their "artificial" independent subsystems, and is qualitatively Nickles nonreducible to that of their main subsystems. However, systems that have topologically different phase space portraits than their unperturbed systems may be integrable singular perturbation problems. If they are, they would be qualitatively different from their *unperturbed systems.* So they are not qualitatively reducible to their unperturbed system in the limit but they can have the same qualitative behavior as the sum of their separable subsystems. For example, the Duffing equation (or the linear damped oscillator) (Nayfeh, 1981, pp.134-136) is integrable (ibid., pp.113-118) but it is a singular perturbation problem. It has no chaotic motion but it still has a topologically different phase portrait from that of the unperturbed system.

Conclusion

In this chapter, we have seen that the behavior of nonseparable systems in classical dynamics is quantitatively and qualitatively different from that of the sum of their "artificial" subsystems (which are independent) and this difference can be defined as the difference between the behavior of dynamically holistic systems and that of dynamically atomistic systems. Dynamically holistic systems are not only nonintegrable but also cannot be approximated by their main subsystems since they are not Nickles dynamically reducible (both quantitatively and qualitatively) to their main subsystems.

That the descriptions of the motions of dynamically holistic systems cannot be Nickles reduced to that of their main subsystems raises problems for theories of exportation which explicitly or implicitly assume that real systems are Nickles reducible to idealized (which usually mean unperturbed) systems. In the next chapter I will show how several current theories of exportation conflict with the Nickles nonreducibility of dynamical holistic systems.

Chapter Three. The conflict between dynamical holism and the current views of scientific methods

Introduction

In Chapter One I have shown that scientific practice itself (outside of QM) breeds a form of holism. This means, according to classical mechanics, that some systems in the world are dynamically holistic and cannot be Nickles reduced to separable systems. That is, the descriptions of the real systems cannot be approximated by the descriptions of idealized systems even when the interactions among the real systems are as small as approaching zero. This raises the problem of exportation: how can we use the descriptions of idealized systems to describe the dynamically holistic real systems? There have been various approaches to tackling this problem. In this chapter I will examine whether these approaches can solve the exportation problem for dynamically holistic systems.

Generally speaking, there are three approaches to solving the exportation problem. The first one can be called the approximation approach: it assumes an approximation relation between the idealized description and the descriptions of real systems to explain why idealized description can be applied to real systems; in short, since the former approximates the latter it is exportable to real cases. However, as we have seen from the previous chapter, the description of an unperturbed system, which is usually taken to be the idealized description, does not approximate that of its dynamically holistic perturbed system since the corresponding perturbation problem is singular. Only when real systems are regular perturbation problems, and thus are Nickles reducible to the unperturbed system, can the description of the unperturbed system approximate that of the real systems. So for the approximation approach to work, real systems have to be regular perturbations of and Nickles reducible to their unperturbed system. In other words, if real systems are dynamically holistic, this approach would not work. McMullin's approximate truth view is an representative of the approximation approach.

The second approach to the exportation problem is to explain exportation in terms of an idealized description's being a part of the description of a real system: since the idealized description is a part of the description of a real system, it can be exported as a description of a part of the real system. This approach requires that we can always identify the idealized description in the description of a real system as an independent constituent of it. A prominent example of this approach is Cartwright's capacity or tendency view. However, this approach also fails in the cases of dynamically holistic systems. As we will see in the next chapter, since dynamically holistic systems have to be treated as singular perturbation problems, their descriptions do not contain the description of the unperturbed system as a constituent. From the last chapter, we can see that only when real systems are regular perturbation problems, can the solution of the unperturbed system be a constituent (i.e., the first term) of the solution of the perturbed system. This means that this approach also requires that real systems be Nickles reducible to their unperturbed systems. In the examination of Cartwright's view, for example, we will see that Cartwright's requirements of an atomistic world and the measurability of tendencies also imply that real systems have to be Nickles reducible to unperturbed systems.

The third approach assumes that what makes the idealized description exportable to real cases is that the idealized system and the real system share the same qualitative behavior. So the idealized description can be exported to describe the qualitative behavior of real systems. But as we have seen in the previous chapter, singular perturbation problems do not have the same qualitative behavior as their unperturbed problems. As I have discussed in Chapter One, dynamically holistic systems have phase space portraits that are qualitatively different from those of the corresponding separable, unperturbed, systems. So this approach also fails in accounting for the exportation to dynamically holistic systems. An example of the third approach is Rueger and Sharp's structural stability view.

Below I will use examples to illustrate each approach and its problems with dynamical holism. I will start with an example of the first approach: McMullin's approximate truth view.

3.1. McMullin's approximate truth view of scientific methods

McMullin thinks our scientific methods are based on Galilean idealizations, the point of which is to approximate the true descriptions of real systems by idealized laws describing simplified, idealized systems. In other words, idealizations have to be "exportable" to approximate the descriptions of real systems. He states, the term "idealization" is

> …to signify a deliberate simplifying of something complicated (a situation, a concept etc.) with a view to achieving at least a partial understanding of that thing. It may involve a distortion of the original or it can simply mean a leaving aside of the some components in a complex in order to focus the better on the remaining ones. The *point* of idealization (but here is the great divide between the Platonic and Aristotelian tradition) is not simply to escape from the intractable irregularity of the real world into the intelligible order of Form, but to make *use* of this order in an attempt to grasp the real world from which the idealization takes its origin." (McMullin, 1985, p.248)

So idealization is not undertaken to get away from the real world and try to describe just the idealized systems but an attempt to describe the real world using the idealized systems. Galilean idealization also implies a process of de-idealization that adds the idealized factors back. The approximate predictions of idealized laws and especially, the improvement of the predictions by de-idealization, justify the view that the idealized laws are approximately true descriptions of the real systems in the world, that is, idealized descriptions are exportable from the idealized systems to the real ones.

McMullin argues the "impediments" eliminated by Galilean idealization (for example, taking the rolling ball as perfectly spherical, taking the small segment of the earth as flat, ignoring the color and material of the ball, eliminating friction and air resistance etc. in the free fall experiment), are either small departures from the truth that can be allowed for, or features that can be added back. Whether the idealization is intended to simplify some features of the real systems or to eliminate the impeding causal factors, the idealizations are undertaken not just to avoid factors that complicate the situation, but with consideration of how these idealizations can be justified and how what has been idealized away can be added back. What needs to be added back into the idealized description should be suggested by the theory itself so that the corrections are not ad hoc. Were the correction to be ad hoc, then even if the corrected theory gave a good approximate fit with the empirical data, the uncorrected idealized theory would still be "false" or "defective." But if the correction is a consistent development of what the theory deliberately left blank for further construction (or filling in), it is a so-called self-correction of the theory. (McMullin, 1985, pp.263-264). McMullin claims this self-correction ability of a theory "is the best testimony of truth" (ibid., p.264), because the "adding back" of the factors that were idealized away improves the theory *only if* "the original model [theory] idealizes the real structure of the object" (ibid., p.261). By "real structure", McMullin means "a set of constituent entities or processes and the relationships between them" (McMullin, 1978, p.139). McMullin explains, "If the original model merely 'saves the appearances' without in any way approximating to the structure of the object whose behavior is under scrutiny, there would be no reason why this reversal of a simplifying assumption, motivated by the belief that the object *does* possess something like the structure attributed to it, would work as it does." (McMullin, 1985, p.262) That is, if the idealized theory does not approximately describe the factors and processes of the real systems, it would be a miracle if its corrected version could make better predictions. So the idealized laws are not only true of the idealized systems, but also justified as approximations to the true descriptions of real systems.

McMullin argues that the Galilean method guarantees the good approximation of the idealized laws to the descriptions of real cases. He says, "The (still counterfactual) claim that all bodies fall at the same speed *in vacuo* was warranted by the way the behavior of bodies of different weights converged on a single simple uniformity as the limiting case of the zero density [of air] is approached." (Ibid., p.267) In other words, the Galilean method assumes that if the behaviors (such as the speeds) of the real or perturbed systems converge to a limit as the perturbation approaches zero, then the limit describes the behavior of the idealized unperturbed system (which is also called the limiting case). Since the speeds of real systems are functions of time, we actually need to require that the convergence is uniform, that is, the convergence holds for all values of *t*.

As we have seen in Chapter Two, only when the solution of an unperturbed equation is a uniform limit, can it be a valid asymptotic approximation to the solution of the equations of the corresponding real perturbed system. If these solutions do not converge uniformly but only pointwise to the solutions of the unperturbed or idealized system, the behavior of the idealized system may not approximate the behavior of the real systems and is not even qualitatively similar to the latter. No matter how small the perturbations are, the behavior of the perturbed systems may still be far away from the behavior of the idealized system. This implies that pointwise convergence (of the perturbed series) would not do the job that McMullin wants it to do. So for McMullin, in order for the descriptions of the idealized cases to be exportable as approximations to the descriptions of the real cases, the Galilean method has to require that the behavior of the real (perturbed) systems converge uniformly to the behavior of the idealized systems. Only under this condition, does this method provide us with the means to justify the idealized laws as true counterfactually or *ceteris paribus* (ibid., p.268).

In the example of free fall, if for some values of *t*, the convergence breaks down, the behavior of real bodies would not warrant the idealized law and the description of idealized system would not approximate the descriptions of real bodies. To guarantee that the idealized law is true and it approximates the descriptions of real systems we have to require that the description of idealized system is the uniform limit of the descriptions of real systems. So the Galilean method not only justifies the correctness of our description of the idealized system, but also justifies the exportation of the idealized description to real systems.

McMullin's requirement that the solutions of the equations of idealized systems are the uniform limits of the solutions of the equations of real systems, as I have shown in Chapter Two, amounts to the requirement that real systems be Nickles reducible to their unperturbed systems. This requirement is obviously not consistent with the suggestion of classical mechanics that the real systems in the world are most likely nonseparable and thus are not Nickles reducible to their unperturbed systems.

From the scientific practice of using singular perturbation methods to deal with nonseparable systems, we can see that the conflict between McMullin's approximate truth account of scientific methods and the dynamical holism of classical mechanics is due to McMullin's view being based on too narrow a view of the perturbation techniques. Science uses not only regular perturbation but also singular perturbation methods. (I will discuss this in detail in Chapter Four). Since dynamically holistic systems are not regular perturbation problems, McMullin's approximate truth view does not apply to them. That is, it cannot account for how we can use idealized models to describe them.

Now I will turn to discuss an example of the second approach: Cartwright's tendency view.

3.2. Cartwright's account of scientific methods

Cartwright interprets McMullin's approximate truth view as an answer only to the question of how we can know what would happen in the ideal case, but not to the question of why we can use idealized laws in real cases or export idealized laws to real cases, which is the core question that an account of scientific methods needs to answer. McMullin's view cannot answer the second question because it supposes that idealized laws approximate the descriptions of real cases but Cartwright thinks idealized laws, which she calls theoretical or fundamental laws "are not true, nor nearly true, nor true for the most part" (Cartwright, 1989, p.175) as descriptions of the systems in the real world. Fundamental laws do not "literally describe the behavior of real material systems." (Ibid., p.203) Even if we add back in all the corrections that the theories can describe, the theories or laws still will never approximate the correct descriptions of the real systems. To get approximate descriptions of the real systems, we have to add ad hoc corrections to the fundamental laws, which then make the laws no longer fundamental (or general) but phenomenological (or concrete). That the fundamental laws do not approximate the real descriptions is especially apparent in modern physics where the effects described by the fundamental laws only make a tiny contribution to the total behavior of the system (ibid., p.188). This indicates that fundamental laws cannot be idealized descriptions that approximate the true descriptions of the real systems.

For Cartwright, to account for the exportation of idealized laws to real cases, we have to see fundamental laws as abstract laws rather than idealized laws. Unlike idealizations which arrange situations to avoid some unwanted factors, abstraction means to "separate the real properties in imagination and ask what does this factor produce by its own nature?" (Ibid., p.198) Abstraction subtracts not only other causal factors, but also the material or context in which the causal factor under study is embedded. So abstract laws are essentially different from idealized laws: while idealized laws describe what would happen when the concrete conditions become ideal, abstract laws describe causal factors that are outside of any context. It is possible to say that the idealized laws approximately describe the behavior of real systems which are under conditions close to the idealized conditions, but it does not make any sense to say abstract laws approximate the descriptions of the real systems. While we can compare a real situation with an idealized situation where all other factors in the real situation are reduced to zero, we cannot compare a real situation that has specific material and concrete features, with an abstract case that does not have any of that. We cannot even compare an abstract case with an idealized situation because idealized situations are also concrete (with material). We can engineer real systems to make them as close to the idealized systems as possible, but we cannot make any idealized or real systems approximate abstract systems.

Cartwright thinks McMullin's mistake is to conflate idealization with abstraction. The Galilean method consists not just of idealizations but of idealizations *and* abstractions. Without abstraction, idealization is useless. Idealizations can only tell us what would happen in ideal conditions. At most, they may explain only the exportation of laws to almost ideal cases. They cannot explain why we can export laws to real cases far from the ideal. Only abstract laws can make sense of such exportations.

(a) Cartwright's tendency view

Cartwright argues that the only way to explain the scientific practice of exporting laws to real cases is through abstract laws. Only the abstract laws can go beyond ideal situations and connect ideal cases with real cases. Since the abstract laws describe the relations between certain factors and their behavior that hold out of context, it implies that these factors always produce those effects in any context. Cartwright calls these stable relations that the abstract laws describe, the *"tendencies"* of the factors to behave in certain ways, to have certain relations with other properties or to produce certain effects. When the factors are causal, she names their associated tendencies, i.e., the causal tendencies to produce certain effects, "*capacities*" (so capacities form a subset of tendencies) (ibid., p.226). Since she seems to think laws describe not only capacities of causal factors but also tendencies of general factors and takes Hamiltonians as descriptions of tendencies (ibid., p.227), I will use the general term "tendencies" below. Cartwright claims that only these stable tendencies can explain why the general laws are exportable to real cases: the effects produced by those factors in abstract (i.e., out of context, rather than merely ideal conditions) will be produced in real conditions. So only abstract laws claiming stable tendencies can be exported to real cases.

For Cartwright, "one looks for what is true in an ideal model in order to establish an abstract law" (ibid., p.191). But while what is true in the ideal model (described by an idealized law) states only what is true in the model and nothing else, the abstract law (which is nothing but the idealized law, but with the ideal conditions dropped for good) can go beyond the ideal conditions and say that the law as a tendency ascription is true in all conditions. (See ibid., pp.191-192) In other words, the abstract law says not only what happens in the ideal conditions but also what effects are produced by these known factors in real conditions no matter what the real manifested effects are. The advantage that the abstract laws have over the idealized laws is due to the fact that the abstract laws are not tied to the occurrent effects or behavior of the factors whereas the idealized laws are. When other factors appear in the real cases, the occurrent effects may be greatly different from the occurrent effects in the ideal cases. The idealized laws describing the occurrent effects of the ideal cases will not be true in the real cases. But the abstract laws as describing the stable tendencies of the factors can still be true in the real cases in the sense that these factors in the real cases still have the same tendencies: they still try to produce the same effects. The only difference between the abstract laws in the ideal cases and in the real cases is that the effects or behavior of the factors described by the abstract laws are manifested in the ideal cases but not in the real cases. In the real cases the effects of these factors described by the abstract laws are contained in the total effects as a part when there are no interactions among the factors.

In summary, Cartwright's tendency view is that fundamental scientific laws are abstract laws and abstract laws imply stable tendencies. She claims the method of Galilean abstraction and idealization "is a method that presupposes tendencies or capacities in nature." (Ibid., p.188) Our scientific practice of exportation of laws without the laws being true or even approximately true of real systems implies that laws describe tendencies. In Cartwright's words, what you learn from an ideal situation tells you something about what will happen in very different, mixed circumstances "*only if* you assume that the factor has a fixed capacity that it carries with it from situation to situation." (Ibid., p.191, my emphasis) This means that scientific methods presuppose an atomistic world which consists of independent "atomic" factors and their separable effects. Scientific methods can only deal with such an atomistic world. We would not know how to go about gaining knowledge if the world were not atomistic. Cartwright thinks that the success of scientific methods in describing real systems in the world indicates that the world is atomistic.

She describes an atomistic world as the following: "Nature selects the capacities that different factors shall have and sets bounds on how they can interplay. Whatever associations occur in nature arise as a consequence of the actions of these more fundamental capacities." (Ibid., p.181) This atomistic world consists of "atoms", each of which exercises its own separate, independent, and invariable effect (ibid., p.156). For Cartwright, these "atoms" are not necessarily microscopic objects; rather, they mean the constituents (such as a harmonic oscillator) of a system at some level of analysis. They can also be the gravitational force and electric force acting on one system. So the Galilean methods do not necessarily reduce systems to the sums of their spatial parts, but generally reduce them to the sums of their atomic factors and reduce the behavior of the systems to the independent tendencies of these atomic factors. This atomistic world obviously is not the dynamically holistic world that classical mechanics assumes. If the world were dynamically holistic, Cartwright's tendency account of scientific methods would be problematic.

(b) The conflict between Cartwright's tendency view and dynamical holism

Cartwright herself considered the possibility of a holistic world (which she called a "wholistic" world). A wholistic world for Cartwright means that factors produce different effects in different circumstances so that the total effects are not the sums of the stable tendencies of these factors but dependent on the whole settings (Cartwright, 1999, p.29). This "would not be a world in which capacities operate." (Ibid., p.157) Even if there are stable phenomena in a holistic world, they do not imply stable tendencies. The stable phenomena could be accidentally produced by the interactions of processes, which only occur in very specific contexts. What we learn about the stable phenomena cannot be generalized or exported to any unknown cases. So in a wholistic world, our scientific knowledge of ideal cases or accidentally stable phenomena would be greatly contextual, which would not help us to describe the real systems. As Cartwright says, if the world is wholistic, that is, "if the commitment to capacities is missing, no further inference can be drawn" (Cartwright, 1989, p.157) and "the hopes to export the laws of the laboratory to the far reaches of the world will be dashed" (Cartwright, 1999, p.31).

However, as I have shown in Chapter One, science itself embraces dynamical holism, which can be seen as a kind of holism in Cartwright's sense. Dynamical holism implies that the essential interactions among the subsystems (which can be seen as the factors or "legal atoms") of a dynamically holistic system make it impossible to reduce the whole system to the sum of the "atoms". Each "atom" (i.e., subsystem) does not have its separable tendencies or behavior independent of other "atoms". This dependence of each subsystem on other subsystems can be seen more clearly from the dependence of the equations of nonseparable systems not only on one pair of variables but also on other pairs of variables (see Chapter One). Recall that the dynamical equations for the gravitational *N*-body problem (which I write again below) are not the equations for the independent subsystems; rather, each equation has to contain variables of more than one subsystem (see Meyer and Hall, p.18):

$$\dot{q}_i = \frac{\partial H}{\partial p_i}$$

$$\dot{p}_i = \frac{\partial H}{\partial q_i}$$

$$H = \sum_{i=1}^{N} \frac{\|p_i\|^2}{2m_i} - \sum_{1 \le i < j \le N} \frac{Gm_i m_j}{\|q_i - q_j\|}$$

So the nonseparable *N*-body system is not atomistic, because each equation involves not only one pair of variables but also the variables $q_1, q_2, \cdots q_N$ and p_i, since the Hamiltonian has a mixture of the variables of the subsystems in the interaction terms. When there are not enough integrals to separate those variables, which is the case for nonseparable systems, the dependence between variables is inevitable. If the whole world is a nonseparable system, Scheibe (1995) argues, strictly speaking we would not be able to apply dynamical laws to any of the systems in the world except to the whole itself. Since the essential interactions among the subsystems of a nonseparable system can be seen as a kind of interaction among factors that change the behavior of the factors and make them dependent on each other, dynamical holism implies that there are no stable dynamical tendencies.

The fault seems to be in Cartwright's assumption of an atomistic world. Although Cartwright herself has worries about the assumption of atomism – as she says, "The more I realized how great is the degree of atomism this world requires, the less plausible it became that this world could be the real world" (Cartwright, 1991, p.4) – she thinks in order to account for the success of the exportation of the idealized laws, we have to assume an atomist ontology (that the world consists of independent factors and their stable tendencies). To account for why laws are successful in making predictions about some systems although they are not even approximately true descriptions of these systems, she assumes that the laws must describe stable tendencies of factors that are manifested in the ideal cases but not manifested in the real cases; nevertheless they still can be used to make effective predictions about real cases. The stable tendencies are not manifested in real cases because of other factors. However, if other factors also have stable tendencies and these stable tendencies are composable (Cartwright, 1989, p.156 and p.227), why can't the composition of tendencies be true, or approximate the description of the real systems? Her composable tendencies assumption seems to contradict her distrust of composition of causes (Cartwright, 1983, p.112). As McMullin points out, "if composition of causes were to be suspect, then causal idealization would automatically be suspect also". (McMullin, 1985, p.273) That is, if composition of causes is not possible, the causal idealization that separates or isolates one cause from the others in order to describe the stable tendency of the isolated cause would not be possible. Cartwright seems to have to assume some kind of wholism, such that the composition or interaction of factors is contextual, in her words, "we usually do not have any uniform procedure for 'adding' interactions" (Cartwright, 1983, p.112), in order to establish the premise that laws, even combination of laws are not true of the real cases to argue for the need of tendencies. On the one hand, she thinks we cannot know how the factors of tendencies would interact from what we know about the tendencies. This is why she thinks we cannot reduce the behavior of a real system to the sum of tendencies but have to add some phenomenological (or context-dependent) interaction terms or modify the tendencies. In this sense she takes the world to be nonreductive or wholistic. On the other hand, to explain why our laws can be so successfully exported to describe the real systems in the world, she thinks the world has to be atomistic, which means that the behavior of these real systems is nothing but the composition of tendencies. If there are always contextual interactions of the tendencies in the real cases, our laws describing stable tendencies would not be so successful in describing them. That Cartwright needs both of the assumptions of a wholistic and an atomistic world for her tendency account implies that there is some inconsistency in her view.

Cartwright might think that we can avoid interactions among the factors in the ideal cases, but that in the real cases there can be contextual interactions, so we can have causal idealization but not composition of causes (in the real cases). Even though most real systems are wholistic in the sense of having contextual interactions of tendencies, what we learn about the tendencies in the ideal cases are still some constituents of the behavior or effects of the real systems. However, this would raise problems: If the interactions of the factors are contextual, the tendencies of the factors would be meaningless. This is illustrated when Cartwright talks about a Hamiltonian which is the sum of a model Hamiltonian plus some correction term (i.e., $H=H_0+H'$). For the model Hamiltonian H_0 to be a tendency, the correction term H' has to be a tendency or a combination of tendencies too. If the correction term is arbitrary or ad hoc, the model Hamiltonian would not be counted as a stable tendency of a certain factor either, for it is always possible to write a Hamiltonian as a sum of some Hamiltonian plus the remainder (i.e., $H=H_0+(H-H_0)$. To assume that there are contextual interactions in the real cases, we would have to assume that the tendencies of the factors interacting in different ways in different real cases have different powers of bringing about their effects or behavior (Cartwright, 1989, p.163). This would undermine the concept of stable tendencies, which assumes the tendencies stay the same or have the same power in all situations.

To reconcile her atomism of stable tendencies with her wholistic view of the world, Cartwright later (for example, 1991, 1999) tries to make the concept of tendencies described in her 1989 book narrower in the sense that they are no more universally stable but vary in different contexts. She considers a holistic world which does not have stable tendencies but has stable phenomena.(Cartwright, 1991) For example, the stable phenomenon of a soliton does not indicate a stable tendency of low-loss optical fiber but is context dependent: it can transmit stable pulses without changing their shapes only when it matches precisely with the input frequencies (Cartwright, 1999, p.30). Although the tendency is context dependent, it is not as highly context dependent as the one in an extreme wholistic world where a law describing the behavior of a factor holds only under a unique specific context and cannot be exported beyond this context at all. The context-dependent "tendency" of low-loss optical fiber can be exported to other cases to the extent that the optical fibers have the right inputs. She calls this view "local realism," which means that the laws are exportable only to the cases where the factors carry the tendencies described by the laws. This local realism commits her to a pluralist view of laws: different domains of phenomena are described by different systems of laws (ibid., p.31) and there are different local tendencies in different domains.

Morrison (1995) points out that Cartwright's local realism still contains a tension. Morrison argues, on the one hand, that for tendencies to account for exporting laws beyond ideal conditions, Cartwright supposes them to be universal except in the cases that involve interactions. On the other hand, for tendencies to be empirical or measurable as Cartwright requires, they have to be measurable in every case in which they hold. This measurability requirement "undermine[s] any argument for the superiority of capacities in virtue of their universality" (Morrison, 1995, p.166) and makes the tendencies no better than *ceteris paribus* laws. Rueger and Sharp (1998) also argue that Cartwright's view of local tendencies introduces *ceteris paribus* conditions in the form of domain or range restrictions of tendencies and would also have the problems of *ceteris paribus* laws. (Rueger and Sharp, 1998, p.7).

Cartwright herself does not seem to take classical mechanics as implying any kind of holism; rather she takes it as the paradigm of her (atomistic) tendencies view. She takes Hamiltonians as tendencies, and therefore assumes that her tendency view can account for the exportations in classical mechanics.

A serious problem with tendencies is that, although the Hamiltonian of a nonseparable system can be written as the sum of a model Hamiltonian and a correction term, since nonseparable systems have to be treated as singular perturbation problems, the solution of their Hamilton's equations cannot be written as the sum of the solution of the corresponding unperturbed problem and the correction terms. The behavior of a dynamical system is described by the solution of its equations of motion, not directly by its Hamiltonian. The Hamilton's equations of nonseparable systems have no exact solutions and when they are treated as perturbation problems in order to get approximate solutions, they turn out to be singular perturbation problems whose solutions are not the sums of the solutions of the unperturbed problems plus some correction terms. The solutions that are written as the sums or functions of the solutions of the unperturbed problems and correction terms, as Cartwright's tendencies view requires, fail to be valid asymptotic approximate solutions of the nonseparable systems. This means that the exportation of the idealized Hamiltonian does not help us to know the dynamical behavior of the dynamically holistic systems.[17] Only when the real systems can be treated as regular perturbation problems, can their solutions be written as sums of the solutions of the idealized Hamiltonians and corrections terms. So Cartwright's tendency view of Hamiltonians implies that descriptions of real systems are Nickles reducible to idealized models and thus her view presupposes a dynamically atomistic world (see Chapter Two).

---

[17] As we will see in Chapter Four, for the perturbation solutions to be valid asymptotic approximations, they have to be written as functions of the terms at different scales. This means that the behavior or solutions of equations of motion of the unperturbed systems cannot be taken as stable tendencies of some atomic factors at some level of analysis, but have to be taken as components or constituent elements at one of the levels or scales of the multiple-level descriptions of nonseparable systems.

(c) The measurability problem of Cartwright's view

As I have explained above, in order to export laws to very non-ideal situations, Cartwright distinguishes abstract laws from idealized laws. While idealized laws are directly related to manifestations of behaviour, abstract laws are not. This makes it possible for abstract laws to be exported as un-manifested stable tendencies to the systems with irregular behavior. Although tendencies usually are not necessarily manifested in behavior, Cartwright insists that tendencies are empirical and measurable. She describes stable tendencies in terms of stable parameters. The tendency of a factor which produces the same effect or behavior with the same strength across the ideal case and the real cases is described by a function of the factor with a fixed parameter. In the simple case in which the effects of different factors can be linearly added up, for example, $x(t) = af(p) + bg(q)$, where the a, b are real numbers and p, q are factors (see Cartwright, 1989, p.165), she requires that the parameters, i.e., the constants a and b, will be the same in the ideal case and the real case. So if a real system contains not only factors p, q but also a factor r, the manifested effect or behavior would be $x(t) = af(p) + bg(q) + ch(r)$.

The measurability of the tendencies depends on the intimate relation between abstract laws and idealized laws: stable tendencies of factors described by abstract laws are revealed in certain ideal circumstances. (Cartwright, 1989, p. 191 and pp.203-204). So although idealized laws describe idealized systems in some contexts (which are eliminated in the abstract laws), abstract laws have to be realized in the ideal cases in order to be measurable. However, this measurability requirement restricts measurable tendencies to the class of systems that can be analyzed as regular perturbation problems. If we are dealing with dynamically holistic systems, tendencies would be hard to measure.

If, as Cartwright indicates, our only epistemic access to tendencies is through ideal cases, in order to know what happens in the ideal cases from our not-perfectly-ideal laboratory cases, we have to require that the behavior of the idealized systems (which means the solutions of the Hamiltonian equations of the idealized unperturbed systems in the case of classical mechanics) are stable against small perturbations. If small perturbations can change the behavior of the idealized systems qualitatively, we would not be able to know what happens in the ideal situations or measure the parameters of the tendencies. This is because although we can, for example, reduce friction and other perturbing factors, we can never make a system really ideal, strictly speaking. There will always be uncontrolled fluctuations in our laboratory systems. So the systems prepared in the laboratory cannot be *exactly* the same as the idealized systems. Only when the idealized system is stable, which means that its behavior is qualitatively similar to the perturbed systems (such as the prepared systems in laboratory), can we come to know the behavior of the idealized system from the behavior of the perturbed systems and there come to know the parameter of the tendency. Rueger and Sharp (1998) take this requirement of stability to be one of structural stability, that is, the topological structure of the behavior of the idealized systems in phase space must be stable against small perturbations. This structural stability would enable us to measure the stable tendencies that are manifested in the idealized systems.

Cartwright seems to think she can measure the stable tendencies by the asymptotic method that McMullin used to justify an idealized law as being true of ideal systems.[18] Like McMullin, she also assumes that the ideal case is a limiting case of the laboratory cases and the behavior of our almost ideal laboratory systems uniformly converges to the behavior of the idealized system. So we can know or measure the ideal cases by extrapolating our laboratory cases to the limit. In terms of perturbation theory, this means that the laboratory systems must be regular perturbations of the idealized systems. If the behavior of the almost ideal cases does not converge to that of the ideal cases uniformly, which means that small perturbations can lead to significant differences between the behavior of the ideal systems and the almost ideal systems, we would not be able to measure the stable tendencies manifested only in the behavior of the ideal systems.

To illustrate this point more clearly, let's revisit some examples used in Chapter Two. First we consider an example of a regular perturbation problem: an oscillator with nearly constant frequency described by the equation:

$$\ddot{x} + (1 + \varepsilon e^{-at})x = 0$$

We can see the term $\varepsilon e^{-at}x$ as describing perturbations of the unperturbed system

---

[18] Cartwright takes McMullin's paper (1985) as providing "solutions to an epistemological problem about how we know what happens in ideal circumstances" (Cartwright, 1989, p.186) However, she thinks exportation is the main problem of scientific methods. So she seems to think McMullin is right on the epistemological problem but that her theory makes new contributions on the problem of exportation.

$$\ddot{x} + x = 0$$

If we apply regular perturbation procedures to the perturbed problem, we will get the approximate solution as the following:

$$x(t,\varepsilon) = \sin t + \varepsilon[\frac{e^{-at}+1}{4+a^2}\sin t + \frac{2(e^{-at}-1)}{a(4+a^2)}\cos t] + O(\varepsilon^2)$$

where the first term *sin t* is the solution of the unperturbed problem. When the perturbation parameter $\varepsilon \to 0$, the solution of the perturbed system uniformly converges to the solution of the ideal system. This enables us to extrapolate what we observe about the behavior of the perturbed system to the limit to know the behavior of the idealized system, that is,

$$\lim_{\varepsilon \to 0} x(t;\varepsilon) = \sin t \quad \text{for all } t$$

But if the perturbed system is not a regular perturbation problem, for instance, the Duffing equation

$$\ddot{x} + x - \varepsilon x^3 = 0$$

where the perturbation term is $-\varepsilon x^3$, the solution of the perturbed system would not uniformly converge to the solution of the idealized unperturbed system. If we apply regular perturbation method to this singularly perturbed system, its solution would be:

$$x(t,\varepsilon) = \sin t - \varepsilon(\frac{9}{32}\sin t - \frac{1}{32}\sin 3t + \frac{3}{8}t\cos t) + O(\varepsilon^2)$$

This solution does not uniformly converge to the solution of the idealized system because the secular term $\frac{3}{8}\varepsilon t\cos t$ becomes the order of the first term (which is the solution of the ideal system) when t is the order of ε^{-1}. So the solution of the ideal case is not the uniform limit of the perturbed case. This means no matter how small the perturbation is, the solution of the perturbed case does not go over into the solution of the ideal case. So we cannot extrapolate what we measure about the behavior of the almost ideal systems to the limit to know the behavior of the ideal case. That is, we cannot measure stable parameters of the tendencies which are manifested only in the ideal cases from our non-ideal experiments.

So the measurability of tendencies requires that real systems are Nickles reducible to their idealized or unperturbed systems. Actually, even the notion of stable tendency itself already implies the requirement of Nickles reducibility of real systems. The notion of stable tendencies requires that each factor have its stable separable effect which can be identified as a constituent in the descriptions of real systems. For example, if the tendency of the factor h_1 is to produce the effect

$f(h_1,a)$ in a real case when there are also other factors such as $h_2,h_3,\cdots$ etc., according to the assumption of stable tendencies, the description of the real case should be something like $f(h_1,a)+l(h_2,b)+m(h_3,c)\cdots$ [19]. But as we have seen in the last chapter, for dynamically holistic systems, we cannot write their solutions as the solutions of their unperturbed systems plus corrections; and as we will see in the next chapter, the solutions of these systems do not contain the solutions of the unperturbed systems as an element. Only when real systems are regular perturbations of unperturbed systems, do the solutions of the real systems contain the solutions of the unperturbed systems as the first term or as an element.

So Cartwright's tendency view presumes a world in which real systems are Nickles reducible to their idealizations and thus cannot solve the exportation problem of dynamical holism.

3.3. Rueger and Sharp's structural stability view of scientific methods

Rueger and Sharp (1998) note the tension between the atomistic world implied by Cartwright's tendency view and the view that the real world is possibly holistic, and try to solve it by their structural stability view. They argue that structurally stable laws can be exported to "tempered" holistic systems and structural stability, while presupposed by Cartwright's tendency account, is alone sufficient to account for the scientific practice of exportation.

---

[19] The form $f(h_1,a)$ is used to indicate that the effect may be a complex (i.e. not simple linear) function of the factor h_1 and the parameter *a*. Even more generally the description could take the form $g[f(h_1,a),l(h_2,b),m(h_3,c)\cdots]$

Rueger and Sharp argue that what scientists presume in exporting laws is not parameter stability, but structural stability, which enables us to export the qualitative (not the quantitative) behavior of ideal system to real systems. Structural stability means that the structure or qualitative behavior of idealized system stays the same under perturbation, or, perturbed systems share the same qualitative behavior with the idealized system. In terms of dynamics, the structure of the behavior of a system is represented by the topology of the phase portrait of the system. In the example we used before:

$$\ddot{x} + (1 - \varepsilon e^{-\alpha t})x = 0\,,$$

if the parameter ε is not stable (that is, its value varies or fluctuates), according to Cartwright's view, we would not able to export the idealized law. However, its "phase space portrait stays qualitatively the same under perturbations such as variations of the value of the parameter" (Rueger and Sharp, p.9). That is, the phase space portrait of

$$\ddot{x} + x = 0$$

is topologically equivalent to that of the perturbed system with a specific value of ε, which according to Rueger and Sharp's view makes the idealized law exportable to the perturbed system. It is not the stable quantitative contribution of a factor in different contexts that makes the law describing it exportable; rather, it is the fact that the qualitative behavior of a configuration of factors stays the same when other factors appear in the real cases that makes the laws describing the stable qualitative behavior exportable. What we export is not the quantitative description of the behavior of the ideal system, but the quality of the behavior of the ideal system. Since structural stability allows parameter variation, it is compatible with a kind of holism that has structural stability but not parameter stability. Extreme forms of holism which do not allow for structurally stable systems would not be hospitable to the exportation of laws. For the account offered by Rueger and Sharp, structural stability is a necessary condition for exportation in virtually all circumstances.

The requirement of structural stability in the context of perturbation theory means that real systems have to be regular perturbations of their unperturbed systems, that is, they must be Nickles reducible to the latter. Most, if not all, singularly perturbed systems are not topologically equivalent to their unperturbed systems. In particular, dynamically holistic systems do not have phase space portraits topologically equivalent to those of their unperturbed (or idealized) systems. As we have seen in Chapter One, the phase portraits of dynamically holistic systems have chaotic regions whereas the phase portraits of their unperturbed systems do not (Lichtenberg and Lieberman, 1983, pp.63-64, and p.41). This means the idealized laws that describe the unperturbed ideal systems are not structurally stable against the perturbations contained in the nonseparable systems. Therefore, structural stability cannot account for the exportation of laws to dynamically holistic systems.

Just as Cartwright tries to use local realism to make stable tendencies locally exportable in a moderately holistic world, Rueger and Sharp think that in some circumstances, exportation can go beyond the limits of structural stability. They claim "even if a model describing solitons is unstable against certain kinds of perturbations, it can often still be used if the theoretical lifetime of the solitons is large against the time scale of the phenomena under investigation" (Rueger and Sharp, pp.14-15). Can these modified views solve the problem of the dynamical holism? It seems not. Nonseparable systems are ubiquitous. This causes difficulties for the very notion and the measurability of tendencies. Although we might be able to find systems that are structurally stable, the exportation from idealized systems to nonseparable systems is not justified since the idealized systems of nonseparable systems are structurally unstable (with respect to nonseparable perturbations). If local structural stability implies that real and idealized systems share similar topologies within some range of the variable, for example, $t < \frac{1}{\varepsilon}$, then this seems to allow real systems to be singular perturbations of (and Nickles nonreducible to) unperturbed systems. However, this local structural stability would also face the *ceteris paribus* problem of restricting idealized laws to very limited applications.

Conclusion

What has been shown? If we have an idealized model, but we don't know the description for the real system, and we try to spell out the conditions under which the application of this model – its "exportation" – to the real system is justified (in the sense of providing a successful approximation of the behavior), we have seen that the theories of McMullin, Cartwright, and Rueger and Sharp fail to provide a justification for the case of nonseparable real systems. In these cases, regular perturbation theory and Nickles reduction break down and I have argued that, in one way or another, McMullin and the others implicitly assume that the relation between idealized models and real systems is to be understood in terms of regular perturbation theory. Dynamically holistic systems cannot be regarded as regular perturbations of idealized models. So the question is: if the theories that require Nickles reduction or regular perturbations cannot account for exportations in cases of dynamically holistic systems, can we have an exportation theory that does not require Nickles reduction?

To construct such a theory we need to find out how our idealized descriptions, in the holistic cases, are related to the descriptions of the real systems. In addressing this question we will again presuppose equations of motion for the real systems and see how we gain approximate solutions to these by applying singular perturbation methods. (Of course, in real cases of exportation, we do not actually have the equations of motion for the real systems at hand). This will lead to a modified notion of idealized models and thus a different view of the process of idealization, i.e., the process through which we find successful (i.e., exportable) idealizations when we know the descriptions of the real systems.

Chapter Four. A solution to the exportation problem in dynamical holism

Introduction

In Chapter Three, I showed that various accounts of scientific practice cannot justify the exportation of an idealized model to dynamically holistic systems because they assume that there is a regular perturbation relation between idealized models and real systems. This relation fails if the real systems are dynamically holistic. Since many real systems are dynamically holistic, we need an account of exportation for such cases. In this chapter, I will show that singular perturbation theory, which provides us with approximate solutions to the real systems, demonstrates new relations between idealized models and real systems that justify exportations of idealized models to dynamically holistic systems. Since singular perturbation theory does not require perturbed systems to be Nickles reducible to unperturbed or idealized systems, this will provide a solution to the exportation problem of dynamical holism.

My strategy in approaching the problem of exportation in dynamically holistic cases is to assume that we (as philosophers) know the equations of motion for the real systems in order to establish the connections between idealized models and real systems. In real cases of exportation, scientists have only the idealized models and typically do not know the equations of motion for the real systems. However, assuming that we do know these equations will enable us to see what kind of approximate solutions we can find for them. These approximate solutions can be seen as exportable idealized models. Since dynamically holistic systems must be treated as singular perturbation problems, we have to use singular perturbation methods to get approximate solutions for them. Singular perturbation methods show that the approximate solutions for dynamically holistic systems are very different from the traditional idealized models, i.e., the solutions of the unperturbed systems. From this we can see that in order for idealized models to be exportable (as approximations) to dynamically holistic systems, we need a modified notion of idealized models and a different view of the process of idealization.

On the other hand, if we are not trying to justify exportations to dynamically holistic systems but to idealize these systems to get successful, i.e., exportable idealized models, the approximate solutions obtained from applying singular perturbation methods are sufficiently different from traditional idealized models so as to suggest that dynamically holistic systems have to be idealized in a distinctive way. We should note that idealization is a different epistemic situation from exportation. To idealize a system means that we are in the situation where we know the equations of motion for the real systems and try to idealize them properly so that the idealized models can be exportable. In contrast, when we are in the *situation of exportation* we know the idealized models but not the equations for the real systems. When we are in the *situation of idealization* of dynamically holistic systems, we need to find approximate solutions to the real systems that can serve as the idealized descriptions. That is, we need to idealize the dynamically holistic systems to the systems that approximate them. We will have to use singular perturbation methods to obtain approximate solutions to the dynamically holistic systems, from which we will see that idealization in dynamically holistic cases has to be done differently than just by taking the solutions of the unperturbed problems as idealized models. This leads to a "new" or modified notion of idealized models.

Although the problem of exporting idealized models to dynamically holistic systems can be solved by introducing the new idealized models that are used in singular perturbation methods, we still need to justify the exportation of traditional idealized models to real cases in order to account for the wide use or universality of idealized laws. If the traditional idealized models cannot be exported as approximations, can they be exported as something else, such as a part of the approximation? If their exportation in this sense can be justified, that would explains why idealized laws can be used universally. Singular perturbation methods show that traditional idealized models do constitute parts of the approximations to the descriptions of a real system. This account of the universality of idealized laws seems similar to Cartwright's tendency account in that both of them explain the universality of idealized laws in terms of the idealized models being a general part of the descriptions of real systems. However, as I will argue, the two are actually very different: while Cartwright's account assumes an atomistic world which conflicts with dynamical holism, the singular perturbation account is consistent with the assumption of a dynamically holistic world but assumes that this holistic world has something like atomistic factors only at certain scales or levels.

Below I will first show how singular perturbation methods obtain approximate solutions to dynamically holistic systems. Then, I will explain why the "new" approximate solutions imply that we need to introduce a modified notion of idealized models, and why these modified idealized models are exportable to the real systems. Finally, I will discuss the question of what traditional idealized models are if they are not approximations and show how singular perturbation methods also provide an answer to this question.

4.1. Singular perturbation methods

Singular perturbation methods are the methods used to get uniformly valid asymptotic approximate solutions to singular perturbation problems. They expand the solutions of singular perturbation problems in ways so that the problem of secular terms would not appear and the uniform asymptoticness of the expansion would not break down. Below I will use an example to show why singular perturbation methods have to introduce multiple scales in the solutions of singular perturbation problems to restore the uniform asymptoticness. The necessity of introducing multiple scales can be seen from the comparison between the exact solution of a singular perturbation problem and the regular expansion of the solution. So to illustrate this we need a problem that has exact solutions, that is, it is integrable but is a singular perturbation problem.[20] The Duffing equation I used in Chapter One is exactly such a problem.

---

[20] Note that although I am using an example that has an exact solution in order to show why regular perturbation methods do not work but singular perturbation methods do work for singular perturbation problems, it does not mean that singular perturbation methods can work only for problems that have exact solutions. As I will show in Section 4.1 (b), singular perturbation methods also work on the problems of dynamically holistic systems, which do not have exact solutions.

(a) An example of a singularly perturbed problem with an integrable Hamiltonian

Recall the Duffing equation (with no damping, describing a conservative or Hamiltonian system):

$$\ddot{x} + x + \varepsilon x^3 = 0$$

with initial conditions at

$$x(0) = 0,$$
$$\dot{x}(0) = a > 0$$

Its exact solution expanded in powers of ε can be written as[21]

$$x(t;\varepsilon) = \sum_{n=1}^{\infty} b_n(\varepsilon) \sin n(t + \varepsilon\omega_1 t + \varepsilon^2\omega_2 t + \ldots)$$

where $b_n = \frac{4}{P}\int_0^{\frac{P}{2}} x(t,\varepsilon)\sin\frac{2n\pi t}{P}dt$

and $P = 2\pi[1 - \frac{3}{8}\varepsilon a^2 + \cdots]$

---

[21] This example is from Kevorkian and Cole, 1996, pp.268-272.

and the ω_i are constants independent of ε. The exact solution shows that it depends on $(t + \omega_1 \varepsilon t + \omega_2 \varepsilon^2 t + \cdots)$ rather than t. Comparing this to the regular expansion

$$x(t;\varepsilon) = a\sin t + \varepsilon a^3 (\frac{3}{8} t\cos t - \frac{9}{32}\sin t - \frac{1}{32}\sin 3t) + \ldots$$

we can see that the problem with the regular expansion is that it takes the solution to be dependent on t alone and therefore it expands the function

$$\sin n(t + \varepsilon\omega_1 t + \varepsilon^2 \omega_2 t + \ldots)$$

as the following:

$$\sin n(t + \varepsilon\omega_1 t + \varepsilon^2 \omega_2 t + \ldots) = \sin nt + n\varepsilon\omega_1 t \cos nt + O(\varepsilon^2) \text{ as } \varepsilon \to 0 \text{ when } t \text{ is fixed}$$

As it is indicated above, the regular expansion is not valid for all t. When t is big, for example, when t is the order of $\frac{1}{\varepsilon}$, the second term in the expansion becomes as big as the order of the first term and breaks the asymptoticness of the expansion. This means that the regular expansion has a singular asymptotic limit, that is,

$$\lim_{\varepsilon \to 0} x(t;\varepsilon) = a\sin t \text{ for fixed small } t$$

But what we want is a solution that is good for both $\varepsilon \to 0$ and $t \to \infty$, that is, a solution that has its first term as its uniform asymptotic limit which guarantees that the solution is uniformly asymptotic

$$\lim_{\varepsilon \to 0, t \to \infty} x(t;\varepsilon) = x_0 \quad \text{for all } t$$

Here x_0 represents the first term of the uniform asymptotic expansion, not necessarily the solution of the unperturbed system.

The difference between the exact solution and the regular expansion is that the exact solution describes a fast change of the phase in terms of the fast time variable t and a slow change of the phase in terms of the slow time variables εt, $\varepsilon^2 t$ etc. These slow time variables $T = \varepsilon t,\ T_1 = \varepsilon^2 t \cdots$ are so called because as $\varepsilon \to 0$, they "condense" the time so that the slow motions can be described over long periods of time. When t is small, the variation of the phase is mainly contributed by the fast change. But when t is big, the slow changes also make significant contributions. By contrast, the regular expansion can only describe the fast change. Instead of describing the slow change in terms of a slow time variable, the regular expansion tries to describe this change in terms of t by the second term $\frac{3}{8}\varepsilon a^3 t \cos t$, which is a secular term. This indicates that to avoid secular terms, a uniform asymptotic approximate solution has to be a function not only of the variable t but also of the variables εt, $\varepsilon^2 t$ etc. That is, $x(t;\varepsilon)$ needs to be written in the form $x(t, \varepsilon, \varepsilon t, \varepsilon^2 t, \ldots)$. One way to introduce the new ε-dependent variables into the solution is to introduce the multiple scales variables $T = \varepsilon t$, $T_1 = \varepsilon^2 t$ etc. in the equation, which is called the method of multiple scales. The introduction of these new independent variables, as I will show below, makes the solution of x_0 undetermined which leaves room for eliminating the secular terms, that is, we can choose the undetermined functions $a(T, T_1 ...)$ and $\beta(T, T_1 ...)$, describing the amplitude and the phase, so that the secular terms will not appear.[22]

The Duffing equation in terms of the multiple scales variables $T = \varepsilon t$, $T_1 = \varepsilon^2 t$ etc. can be written as

$$\frac{\partial^2 x}{\partial t^2} + 2\varepsilon \frac{\partial^2 x}{\partial t \partial T} + \varepsilon^2 (2 \frac{\partial^2 x}{\partial t \partial T_1} + \frac{\partial^2 x}{\partial T^2}) + x + \varepsilon x^3 + ... = 0 .$$

x is then expanded as

$$x = x_0(t, T, T_1 ...) + \varepsilon x_1(t, T, T_1 ...) + ...$$

---

[22] Since in the example we introduced multiple time scale variables only in the form $T = \varepsilon t, T_1 = \varepsilon^2 t \cdots$, it may create the false impression that multiple scales have to be in this form. But the procedure of multiple scales can be generalized to include the use of other forms of multiple scales such as $\frac{t}{\varepsilon}, \frac{t}{\sqrt{\varepsilon}}, etc.$ (see Holmes, 1999, p.32).

The notion of multiple scales may also be extended to include spatial scales to get asymptotic approximate solutions to the problems described in terms of spatial variables. For example, to use the multiple scales method to get the asymptotic approximate solution of the following equation describing steady state heat conduction in a one-dimensional rod (see Hinch, 1990, pp.126f.)

$$\frac{d}{dx}[(k(x) \frac{dT(x)}{dx}] = Q(x)$$

we need to introduce two length scales: the microscopic position variable $\xi = \frac{x}{\varepsilon}$, and the macroscopic position variable x to describe the motions at macroscopic and microscopic levels.

Inserting this expansion into the Duffing equation and comparing the coefficients of ε^n we get

At ε^0: $\frac{\partial^2 x_0}{\partial t^2} + x_0 = 0$

The general solution of the above equation can be written as:

$$x_0 = a(T, T_1 \ldots)\cos[t + \beta(T, T_1 \ldots)]$$

(Comparing this to the exact solution, the β function in the solution of the x_0 above representing a correction in the phase, is a function of $\omega_1, \omega_2, \cdots \omega_n; T, T_1, \cdots T_n$, or more specifically, $\beta(T, T_1 \cdots) = \omega_1 T + \omega_2 T_1 + \cdots$.)

From this solution we can see that because of the introduction of the multiple scales variables, the amplitude and frequency of the first term of the expansion of x are no longer constants (as in the case of the regular expansion) but functions of the slow time scale variables, T, T_1, ... They are undetermined functions so we get the freedom to impose more conditions to make the solution unique, which makes it possible to prevent the secular terms from appearing by choosing these conditions to be such that the secular terms would be eliminated. In order to determine $a(T, T_1 \ldots), \beta(T, T_1 \ldots)$, we need to proceed to the next level of approximation, that is, the level of $O(\varepsilon)$.

Comparing the coefficients of the terms at ε^1 or $O(\varepsilon)$ we get

$$\frac{\partial^2 x_1}{\partial t^2} + x_1 = -2\frac{\partial^2 x_0}{\partial t \partial T} - x_0{}^3$$

Putting the above solution of x_0 into this equation we obtain

$$\frac{\partial^2 x_1}{\partial t^2} + x_1 = 2\frac{\partial a}{\partial T}\sin(t+\beta) + (2a\frac{\partial \beta}{\partial T} - \frac{3}{4}a^3)\cos(t+\beta) - \frac{1}{4}a^3\cos(3t+3\beta)$$

Upon integration, the first and second terms on the right-hand side produce secular terms. However, since now these terms contain undetermined a, β, we can make these terms disappear by choosing a, β appropriately. Thus we have the so-called solubility or secularity conditions,

$$\frac{\partial a}{\partial T} = 0$$

whose solution is $a = a(T_1, T_2 ...)$; and

$$2a\frac{\partial \beta}{\partial T} - \frac{3}{4}a^3 = 0$$

whose solution is

$$\beta = \frac{3}{8}a^2 T + \beta_0(T_1, T_2 ...)$$

If we are interested only in a solution that is valid for $t = O(\varepsilon^{-1})$, we can take a and β_0 to be constants since they are functions of only T_1, T_2 etc. Then the solution can be written as

$$x = a\cos(t + \frac{3}{8}\varepsilon t a^2 + \beta_0) + O(\varepsilon)$$

This solution is uniformly valid for all times up to $O(\varepsilon^{-1})$.

From the solution obtained from the singular perturbation method we can see the following new features compared to the solutions of regular perturbation problems:

(i) The solution, especially the first order terms,[23] contains multiple scales. We can see from the solution of the Duffing equation that the first order term $a\cos(t + \frac{3}{8}\varepsilon t a^2 + \beta_0)$ in the solution contains not only the variable t but also a new slow time variable $T = \varepsilon t$. Because of the introduction of the slow time variable T, the solution becomes valid for a large time domain: $t < \frac{1}{\varepsilon}$. We can also introduce slower time scales such as $\varepsilon^2 t$ to describe the slower motion over a longer period of time. The first term then would also contain this slower time scale variable which would make the solution valid for a larger time domain: $t < \frac{1}{\varepsilon^2}$.

---

[23] The "first order terms" means the terms of the first order, that is the order of $O(1)$. In some cases, there is one such term in the solution; in other cases, the first two or more terms in the solution may be first order.

This means that the multiple scales method implies that x varies at different time scales and we cannot reduce these different processes at different time scales to a process at one scale. That the leading term is not just a function of t as in the regular expansion, but a function of t and other time scales, indicates that there are motions at different time scales at the order of the leading term. In the example of the Duffing equation, the leading term of the solution is a periodic function with a constant amplitude. The change of the phase contains several simultaneous processes. We can see this more clearly if we write the solution of the Duffing equation in the general form as:

$$x(t;\varepsilon) = a(T, T_1 \cdots)\cos(\omega_0 t + \omega_1 T + \cdots)$$

The term $\omega_0 t$ describes a fast change in the phase at the time scale t with the frequency ω_0 (which is 1 in the previous example). The term $\varepsilon\omega_1 t$ ($\omega_1 = \frac{3}{8}a^2$ in the example) describes a slow change of the phase at the time scale εt with the frequency ω_1. This slow motion at the time scale εt is superimposed on the fast motion at the time scale t. When t is small, the phase change is mainly contributed by $\omega_0 t$. The other terms describing slow motions in the phase are just small corrections. But when t is as large as $\frac{1}{\varepsilon}$, the second term $\varepsilon\omega_1 t$ in the phase becomes the order of $O(1)$ and cannot be ignored. So we need to keep terms up to $\varepsilon^n \omega_n t$ in order to make the expansion uniformly valid for t of the order of $\frac{1}{\varepsilon^n}$. This implies that the uniform expansion has the leading term as a compound limit in terms of several independent variables.

In contrast, the leading term of the regular expansion, $a\sin t$, can only describe the motion at the fast time scale t. This is because the regular expansion treats the effect of perturbation terms as of order ε or higher, so the perturbation has no effect on the leading term and therefore the leading term is just the solution of the unperturbed problem. But for singular perturbation problems, the cumulative effect of perturbation terms can be $O(1)$ when t is big enough (such as $t = O(\varepsilon^{-1})$ or bigger) so that the effect of the perturbation becomes part of the leading term rather than just being in the correction terms. But the effects of the perturbation are the motions at slow time scales. The problem with the regular expansion is that it tries to describe these motions at slow time scales by the fast time variable t. While the frequency of the exact solution depends on ε and nonlinearly depends on the amplitude, the first term of the regular expansion has to force the frequency to unity. To describe the slow changes in the phase with the independent variable t, the regular expansion has to introduce the second term, i.e., the secular term. So the attempt to describe the motions at slow time scales by only the fast scale variable breaks the asymptoticness of the expansion. In other words, the attempt to reduce the description to one scale fails.

(ii) The solution of the unperturbed system is a "one scale part" of the first order terms. If we look at the motion described by the first order terms only at the fast time scale t (in the example of the Duffing equation), that is, if we let $\varepsilon t \to 0; \varepsilon^2 t \to 0 \cdots$, then we get $x(t) = a\cos(t + \beta_0)$, which is the solution of the idealized unperturbed system.[24] This indicates that the solution of the unperturbed system constitutes a "one scale part" of the first term $a\cos(t + \frac{3}{8}\varepsilon t a^2 + \beta_0)$ and describes the motion of the perturbed system at the fast time scale t.

(iii) The first order terms include the effect of the average part of the perturbation. As explained in the last paragraph, the first order term is not the solution of the idealized unperturbed system but contains the latter as a part. What else does it contain besides the solution of the idealized unperturbed system? We may guess that the other part of the first order terms describes some effect of the perturbation. To find out what effect of the perturbation the term contains, we need to look at the procedure of the averaging method, which produces the same first term as the multiple scale method, but shows more clearly where the other part of the term comes from.

To use the averaging method we first need to transform the second order Duffing equation into two first order equations. By taking the a (the amplitude) and β (in the phase) in the general solution of the ideal system as variables, we can transform the Duffing equation into two first order equations in terms of these two variables.[25] In these two equations there are slowly varying perturbation terms and fast varying perturbation terms. By taking the average of the perturbation terms over the interval $[0, \pi]$, the oscillation terms will disappear since their averages are zero. We are left with only the slowly varying term, which is the (non-zero) average of the perturbation term. So we replace the perturbed system with the so called averaged system with only the average term(s) and solve the averaged problem. If we transform the solution of the averaged system back to expressions involving the old variable x, we get exactly the same solution as we got from the multiple scales method. So the averaging method, as Arnold puts it, simply means ignoring the purely periodic term in the first order equations. Although the periodic terms in the first order equations have the same order as the other terms, they oscillate and lead only to small oscillations of the solution. The non-oscillation terms, i.e., the slowly varying terms, lead to a drift which over a long time period (e.g., $\frac{1}{\varepsilon}$) modifies the solution of the idealized unperturbed system by a quantity of order 1 (Arnold, 1983, p.140).

---

[24] Recall the unperturbed or the ideal system of the Duffing equation is $\ddot{x} + x = 0$. The general solution of it is $x = a\cos(t + \beta_0)$.

[25] The procedure of solving the Duffing equation by the averaging method can be found in Nayfeh, 1981, pp.127-131.

The averaging method shows that the first term of the singular asymptotic approximate solution contains not only the solution of the ideal system but also the effect of the average part of the perturbation.

(b) Solving the equations of dynamically holistic systems by singular perturbation methods

The Duffing equation used above has an exact solution, which means that it describes an integrable or separable system. Some may wonder whether this multiple-scales technique also works for nonseparable problems of many degrees of freedom (so that we can get solutions for dynamically holistic systems). To show that the multiple scales method can also solve such problems such as the many-body problem, I will first show that this method can solve nonseparable problems of one degree of freedom.[26] Since all Hamiltonian problems of one degree of freedom are separable, I will have to use a non-Hamiltonian problem of one degree of freedom as an example. Then I will show how the procedure can be extended to solve the Hamiltonian problems of many degrees of freedom. We may also take solving nonseparable non-Hamiltonian problems of one degree of freedom as a way of solving nonseparable Hamiltonian problems of many degrees of freedom since the (interacting) subsystems of the Hamiltonian systems of many degrees of freedom can be seen as nonseparable non-Hamiltonian systems of one degree of freedom.

A simple example of a nonseparable non-Hamiltonian problem of one degree of freedom is the damped oscillator described by the van der Pol equation:[27]

$$\ddot{x} + x + \varepsilon\dot{x}(x^2 - 1) = 0\,,$$

with initial conditions

$$x(0) = a_0 \quad \dot{x}(0) = 0$$

---

[26] There are several other singular perturbation methods besides the multiple scale method. The Lindstedt-Poincaré method and renormalization method are effective only for problems with periodic solutions [see Nayfeh ((1973) p.99 and (1981) p.152] whereas the multiple scale and averaging methods are effective for general problems with varying amplitude solutions. But the averaging method gets the same results as the multiple scale method and all the singular perturbation methods introduce new variables at different scales in the solutions. For the purpose of my thesis, I will introduce only the multiple scale method here.

[27] This example is from Nayfeh (1981), Chapter 6.

Applying the multiple scales method to the van der Pol equation (which includes introducing the slow time variable $T = \varepsilon t$), we get the uniform asymptotic solution to the first approximation as:

$$x(t,T) = 2(1 + (\frac{4}{a_0^{\ 2}} - 1)e^{-T})^{-\frac{1}{2}} \cos t + O(\varepsilon),$$

which implies that for $t \to \infty$

$$x \to 2\cos t + O(\varepsilon)$$

This indicates that the solution has a limit cycle, that is, no matter what the value of the initial condition a_0 is, as long as it is not zero, the system will tend towards the same periodic behavior. In contrast, the solution of the unperturbed problem is

$$x_0(t) = a_0 \cos t$$

which describes a periodic motion determined by the initial conditions, not a limit cycle motion. The amplitude of the uniform asymptotic solution increases at the time scale of $T = \varepsilon t$ whereas the first term of the regular expansion which is the solution of the unperturbed problem does not.

The asymptotic solution of the van der Pol equation also contains two processes, namely, an oscillation at the fast time scale t and a drift of the amplitude at the slow scale T. Compared to the solution of the Duffing equation which has the slow motion in its phase, the solution of the van der Pol equation has the slow motion in its amplitude, which can be seen clearly from Figure 4. In Figure 4, the continuous curve describes the fast oscillation (with drifted amplitude) and the dashed curves represent the drift of the amplitude. So the asymptotic solution of the van der Pol equation is a combination of two simultaneous processes at different time scales: fast oscillation (without drifted amplitude) at the fast time scale t and the slow drift of amplitude at the slow time scale T. When t is big, the slow motion modifies the fast oscillation significantly. The fast oscillation without the drift of the amplitude is described by the solution of the unperturbed system. So the solution of the unperturbed system constitutes one of the processes of the perturbed system at one time scale.

The multiple scales method can also be applied to approximately solve equations of nonseparable Hamiltonian systems of many degrees of freedom, e.g., the many-body problem. The many-body problem as a problem of N-bodies (e.g., the sun and the planets) with gravitational interactions can be described by the following equation (see Meyer and Hall, 1992, pp.17-18):

$$m_i \ddot{x}_i = \sum_{j=1}^{N} \frac{G m_i m_j (x_j - x_i)}{\|x_j - x_i\|^3} \qquad i = 1, 2, \cdots N$$

where G is the gravitational constant and m_i is the mass of the i^{th} body.

The above equation can be transformed into the following first order equations:

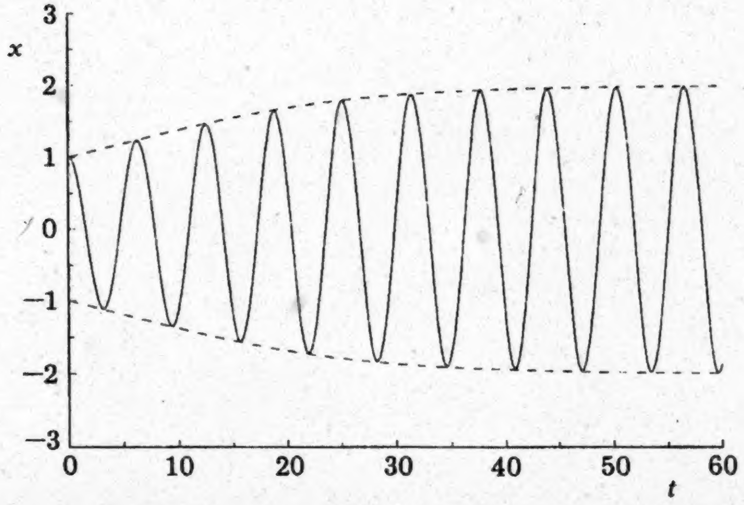

Fig. 7.1 The continuous curve is a numerical solution $x(t;\epsilon)$ of the initial value problem for a van der Pol oscillator with $\epsilon = 0.1$. The dashed curves give the asymptotic predictions for the amplitude, i.e. $\pm R(T)$. The agreement is good at even larger values of the small parameter, as is typical of this class of asymptotic analysis.

Figure 4

From E. J. Hinch, 1981, P.109, figure 7.1

$$\dot{q}_i = \dot{x}_i = \frac{p_i}{m_i} = \frac{\partial H}{\partial p_i}$$

$$\dot{p}_i = \sum_{i=1}^{N} \frac{Gm_i m_j (q_j - q_i)}{\|q_i - q_j\|^3} = -\frac{\partial H}{\partial q_i}$$

$$H = \sum_{i=1}^{N} \frac{\|p_i\|^2}{2m_i} - \sum_{1 \le i < j \le N} \frac{Gm_i m_j}{\|q_i - q_j\|}$$

The *N*-body problem, can be seen as a perturbation of the unperturbed system describing the motion of *N*-1 noninteracting planets moving around the Sun which is at rest. So the unperturbed system can be decomposed into *N*-1 Kepler problems. The perturbations would be the small gravitational interactions among the planets (see Arnold, 1983, p.139 and p.171). The *N*-body problem can be described by the *N*-1 pairs of first order equations in the standard form

$$\frac{dp_n}{dt} = -\frac{\partial H}{\partial q_n}$$

$$\frac{dq_n}{dt} = \frac{\partial H}{\partial p_n}$$

$$H = H(p_1, \cdots p_{N-1}, q_1, \cdots q_{N-1}; \varepsilon) = [\sum_{i=1}^{N-1} H_i(p_i)] + \varepsilon H_{\text{int}}(p_1, \cdots p_{N-1}, q_1, \cdots q_{N-1})$$

$$n = 1, 2 \cdots N\text{-}1$$

where $\sum_{i=1}^{N-1} H_i(p_i)$ is the sum of the Hamiltonians of the N-1 Kepler problems, $\varepsilon H_{\text{int}}(p_1,\cdots p_{N-1},q_1,\cdots q_{N-1})$ is the sum of the Hamiltonians of the interactions among the planets and H is 2π-periodic in the q_i (see Arnold, 1983, p.168 and p.171).

If we decompose H into its average $\overline{H}$ and oscillatory $\hat{H}$ parts and expand it in powers of ε, we have

$$H=\sum_{i=1}^{N-1}(\overline{H}_0)_i(p_i)+\varepsilon\sum_{i=1}^{N-1}(\overline{H}_1)_i(p_i)+\varepsilon\hat{H}_1(p_1,\cdots p_{N-1},q_1,\cdots q_{N-1})+\varepsilon^2\sum_{i=1}^{N-1}(\overline{H}_2)_i(p_i)$$
$$+\varepsilon^2\hat{H}_2(p_1,\cdots p_{N-1},q_1,\cdots q_{N-1})+O(\varepsilon^3)$$

Then the N-1 pairs of first order equations can be written as (see Kevorkian and Cole, 1996, pp.440-441 and pp.386-387)

$$\frac{dp_n}{dt}=-\varepsilon\frac{\partial\hat{H}_1}{\partial q_n}-\varepsilon^2\frac{\partial\hat{H}_2}{\partial q_n}+O(\varepsilon^3)$$

$$\frac{dq_n}{dt}=\frac{\partial H_0}{\partial p_n}+\varepsilon(\frac{\partial\overline{H}_1}{\partial p_n}+\frac{\partial\hat{H}_1}{\partial p_n})+\varepsilon^2(\frac{\partial\overline{H}_2}{\partial p_n}+\frac{\partial\hat{H}_2}{\partial p_n})+O(\varepsilon^3)$$

If we apply the method of multiple scales to the above equations, we introduce the slow time scale variable $T=\varepsilon t$ and expand p_n, q_n in powers of ε and also expand the averaged and oscillatory functions in the above equations. Then we compare the terms of the same order and get equations at orders $O(\varepsilon)$, and $O(\varepsilon^2)$. By using the secularity or solubility conditions we can get the uniform asymptotic approximate solutions for p_n, q_n to first approximation as the following

$$p_n(t,\varepsilon) = a_n + O(\varepsilon)$$

$$q_n(t,\varepsilon) = \tau_n + \eta_n(T,\beta_i) + O(\varepsilon)$$

where a_n are the constants of integration from solving the unperturbed equation $\dot{p}_n = 0$; τ_n are the fast changes in phase $\tau_n = \frac{1}{\varepsilon}\int_0^T \omega_n ds = \omega_n t$, which is described in terms of the fast time scale variable t (and is the solution for the q_n of the unperturbed system); and η_n are the slowly varying phase shifts which are described in terms of the slow time scale variable T.[28]

By comparison, the solution of the unperturbed problem:

$$H = \sum_{i=1}^{N-1} H_i(p_i)$$
$$\dot{p}_n = -\frac{\partial H}{\partial q_n} = 0$$
$$\dot{q}_n = \frac{\partial H}{\partial p_n} = \omega_n$$

is

---

[28] η_n can be determined by solving the equations resulting from secularity conditions (see Kevorkian and Cole, 1996, pp.395-403).

$$p_n = a_n$$
$$q_n = \int_0^t \omega_n d\sigma = \omega_n t$$

From the solution we can see that:

(a) The solution of the N-body problem contains two time scales of variables, namely, the fast time variable t and the slow time variable $T = \varepsilon t$.

(b) The solution of the N-body problem contains the solution of the unperturbed system as a "one scale part", that is, when we set the slow time scale variable to zero, we get the solution of the unperturbed system. In terms of the variables p_n, q_n, the solution of the N-body problem $p_n(t,\varepsilon) = a_n + O(\varepsilon)$ and $q_n(t,\varepsilon) = \tau_n + \eta_n(T, \beta_i) + O(\varepsilon)$, seems to be the solution of the unperturbed system plus correction terms. This may cause us to think that the solution in this form can be used to restore the theories of McMullin and others. However, the correction term η_n is not small. This means that the solution of the unperturbed system is not the uniform asymptotic limit of the solution of the perturbed system in terms of the variables p_n, q_n. By way of another example, if we consider a system of weakly coupled nearly linear oscillators and transform the solutions of the perturbed system in terms of the variables p_n, q_n back into the solution in terms of the variables x_n, the amplitude of the x_n would be some function of p_n and the variation in the phase of x_n would be described in terms of q_n (see Kevorkian and Cole (1996), pp.388-389). So the solution of the perturbed system in terms of x_n would not be in the form of the solution of the unperturbed system plus corrections.

(c) Although I did not show how the η_n are determined in detail, the fact that they are determined from the secularity equations that involve only average terms of the perturbation indicates that they are the effects of the average part of the perturbation (see Kevorkian and Cole, 1996, p.402).

(d) The solution for each p_n, q_n depends on t, ω_n only. This means that the approximate solution of the N-body problem becomes the approximate solutions for the subsystems, each of which is described by a pair p_n, q_n. This is not a result forced by our assumption that each p_n, q_n can be expanded in powers of ε; as Kevorkian and Cole show, it is a consequence of the secularity conditions (ibid., p.398).

We should note that the above approximate solution fails at resonance, that is, when the frequencies ω_n are so related that there are zero divisors in the solution. (I will discuss the small divisor problem in more detail below). This means that the method of multiple scales (as well as the method of averaging) is valid only under the condition that no resonances are possible in the solution to the order of $O(\varepsilon^{n+1})$ if we are interested in the solution to the order of $O(\varepsilon^n)$ (ibid., p.403).

4.2. The conditions for the asymptotic solutions of singular perturbation methods to be valid

Mark Holmes (1999) shows that for a general differential equation of the following form

$\ddot{x} + \varepsilon f(x,\dot{x}) + x = 0$ for $t>0$

where $x(0) = a, \dot{x}(0) = b$

if $f(x,\dot{x})$ is polynomial, then the multiple scale expansion $\sum_{i=0}^{N} x_i(t,\varepsilon t)\varepsilon^i$ is uniformly valid for $0 \le t \le \frac{T}{\varepsilon}$ for fixed $T>0$.

However, as Kevorkian and Cole indicate, in the asymptotic solutions to the problems with many degrees of freedom obtained from the multiple scale method or averaging method, there is always the possibility of zero divisors which invalidate the asymptotic expansion, that is, the small divisors break down the asymptoticness of the solution (Kevorkian and Cole, 1996, p.469, also p.403). In the example of the many-body problem in the last section, using the multiple scales method we need to integrate the oscillatory term, which would introduce small divisors in the solution.

Poincaré called the small divisor problem the "fundamental problem" of classical mechanics. Tabor described the small divisor problem as one that has "plagued advances in classical mechanics over the past hundred years" (Tabor, 1989, p.104). Some held that the small divisor problem implies that the addition of even the smallest nonintegrable perturbation would make the systems ergodic, which means they would have no regular trajectories (see Tabor, p.105, Lichtenberg and Lieberman, 1983, p.51).

The small divisor problem is solved by the so called KAM (Kolmogorov-Arnold-Moser) theory. The KAM theory developed a "superconvergent" method to overcome the divergence of the asymptotic expansion of the multiple scale methods. The "superconvergent" method is a development of the averaging method, which chooses the canonical transformations so that "the order of the perturbation is increased by the square of the preceding one for each step" (Lichtenberg and Lieberman, 1983, p.139), that is, the *n*th step of transformation pushes the oscillatory terms to the order of ε^{2^n}, which "counteracts the influence of the small denominators" (Arnold, 1983, p.181) and makes the expansion convergent (Lichtenberg and Lieberman, 1983, p.143). Similarly, the KAM expansion also contains multiple scales. (Ibid., pp.141-143)

The KAM superconvergent method provides the basis for the following KAM theorem:

If the unperturbed system is nondegenerate, (which means that $\det\left|\frac{\partial \overline{\omega}_i}{\partial p_j}\right| = \det\left|\frac{\partial^2 H_0}{\partial p_i \partial p_j}\right| \neq 0$) and if the perturbation is small enough, most nonresonant invariant tori of the unperturbed system do not vanish but are only slightly deformed (Arnold, 1983, p.183, or Tabor, 1989, pp.105-106).

The KAM theorem proves that the motion of singularly perturbed systems is mostly (i.e., for most initial conditions) regular. This shows the uniform asymptotic solution, which describes the solution of the perturbed system as a modification of the solution of the unperturbed system, to be generally correct. So when the asymptotic solution does not blow up because of small divisors, that is, when there are no resonances possible in the solution up to the order of $O(\varepsilon)$ or higher, the asymptotic solution resulting from multiple scales or averaging methods is a valid approximations to the first order or higher (see Kevorkian and Cole, 1996, p.403 and p.469).

So the asymptotic solutions are justified as valid approximations of the solutions of the real systems, under certain restrictions. They are valid (1) for systems whose unperturbed systems are nondegenerate and their perturbations are small enough (as indicated by the KAM theorem); (2) for most initial conditions of these systems described in condition (1) (as the KAM theorem implies); (3) to the order of ε^n with the error of the order ε^{n+1} if we truncate the expansion at the terms of order ε^n, (see Arnold, 1988, p.150) (4) for a wide domain of the variable (for example, if we introduce the multiple scales variables : $T = \varepsilon t, T_1 = \varepsilon^2 t, ... T_{n-1} = \varepsilon^n t$, the approximate solutions are valid for the domain of the variable *t*: $0 \le t \le \varepsilon^{-n}$ (see Holmes, 1999, p.32)).

4.3. An account of idealization and exportation in a dynamically holistic world

From the above illustration of the singular perturbation methods we can see that a special characteristic of singular perturbation methods is that they introduce multiple scales in the solutions of singular perturbation problems. Below I will explain why it is necessary to introduce multiple scales to dynamically holistic systems to obtain approximate solutions for them, and how this necessity implies that these systems need to be idealized differently than in the traditional way of idealization, which takes the solutions of the unperturbed systems as idealized models. I will then show how the multiple scales approximate solutions of real systems provide a means for justifying exportation in dynamically holistic cases.

(a) Multiple scales are necessary for obtaining approximate solutions for the dynamically holistic systems

Singular perturbation methods show that the way to get approximate solutions of dynamically holistic systems is to impose multiple scales on them, that is, to describe them in terms of multiple scales variables. It is by imposing the multiple scales variables on dynamically holistic systems that uniform asymptotic approximations get restored. While the solution of a dynamically holistic system described in terms of a variable at one scale cannot be divided into an idealized unperturbed part (the solution of the idealized unperturbed system) and a small perturbation part since the effect of the perturbation contains secular terms, the introduction of the multiple scales makes it possible to "divide" a dynamically holistic system in a new way, that is, to divide it into an integrable part and a small (nonintegrable) correction part so that the solution to its equation of motion is the solution to the equation for the integrable part plus small corrections. It is this new way of dividing a dynamically holistic system that makes the asymptotic approximation of its solution possible: by truncating the small corrections in the solution, we get the approximate solution from the solution for the integrable part. Although the solution for the integrable part is, to the first approximation, the uniform asymptotic limit of the solution to the dynamically holistic problem, the integrable part is not the solution for the unperturbed system but for the unperturbed system plus the average part of the perturbation. In other words, the small correction part is not the total effect of the perturbation but only the effect of the small oscillatory part of the perturbation.

The restoration of the uniform asymptoticness by multiple scales does not mean that the dynamically holistic systems are Nickles reducible to the integrable systems described by the uniform asymptotic solutions. This is because the solution of the integrable part to the first approximation is still not the uniform asymptotic limit. Although theoretically we could introduce more slow time scale variables to make the approximate solution valid for the time variable t as big as we want, it is still not valid for all t. As the solution is not valid for all t, there would still be the possibility of the breakdown of the asymptoticness for some values of t.

However, the integrable parts described by the uniform asymptotic solutions are the sum of the motions of their separable subsystems which are described in terms of multiple scales variables. So the introduction of the multiple scales variables also makes the solutions of the subsystems divisible into integrable parts and small correction parts so that the approximate solution of the whole system can be reduced to the approximate solutions of the subsystems. This means that by describing the dynamically holistic systems in terms of multiple scales variables we can obtain approximate solutions of them from the solutions of their integrable parts, i.e. from the approximate solutions of their subsystems.

(b) Exportations and idealizations in dynamically holistic cases

What new features, then, does the introduction of multiple scales in the solutions of singular perturbation methods imply about exportation? That is, how can idealized models be exported to real systems; how are idealized models connected to real systems in the light of singular perturbation methods? First, the necessity of introducing multiple scales in the approximate solutions of dynamically holistic systems implies that the relation between idealized models and real systems cannot be understood in terms of regular perturbation theory, as assumed by the theories of McMullin, Cartwright, and Rueger and Sharp. We have seen that in the case of dynamically holistic systems, assuming regular perturbation theory as the way of connecting idealized models and real systems leads to a problem: the behaviors of unperturbed and perturbed systems differ in ways that do not allow us to regard the unperturbed behavior as a (qualitative) approximation of the perturbed behavior. Hence, "exportation" of the description of the behavior from the idealized model to the real system is not justified. Simply adding back in more perturbations into the idealized models does not solve this problem because that does not help to approximate the behavior of the real system. This is why the theories of McMullin, Cartwright, and Rueger and Sharp cannot justify exportations in holistic cases.

Second, the necessity of introducing multiple scales to gain an approximate description of the behavior of the real system implies that we need to idealize the real system differently than taking the solution to the unperturbed problem as the idealized model, in order to find exportable idealized models. As we have seen, treating the perturbed system according to singular perturbation methods shows how to achieve an approximate description of the behavior of the real system. But these approximations now look different from the ones we would get from regular perturbation methods. Thus, if a system is dynamically holistic, a successful (i.e., exportable) idealized model will in general have to look different from the models we get by just taking the solutions of the unperturbed system. Only for these "modified" idealized models does exportation (in the sense of providing valid approximate solutions) make sense.

In the dynamically holistic case, the process of idealization itself, therefore, has to be different from the received or traditional view of idealization. The traditional notion of an idealized model (such as Cartwright's and McMullin's) implies that the model has all the perturbing factors removed. In terms of perturbation theory, it would be the unperturbed system described by the equation of motion without perturbation terms. This means to idealize the description of the behavior (i.e., the solution to the equation of motion) of the real system is to idealize its equation of motion. This notion of an idealized model assumes stable idealization, that is, we will get the same result whether we first idealize the perturbed equation to the unperturbed equation (by letting the perturbation parameter in the equation go to zero) and then solve the idealized equation, or whether we first solve the exact equation and then idealize the solution (by letting the perturbation parameter in the solution go to zero).

The requirement of the commutation of the operations of solving and approximating equations can be illustrated by the diagram below[29]

$$\begin{tikzcd}
\mathit{equation}(\varepsilon = 0) \arrow[r] & \mathit{solution}(\varepsilon = 0) \\
\mathit{equation}(\varepsilon \neq 0) \arrow[r] \arrow[u] & \mathit{solution}(\varepsilon \neq 0) \arrow[u]
\end{tikzcd}$$

The results of traditional idealizations are the equations and solutions of the unperturbed problems (which are described in terms of only one scale variable). But as I argued in Chapter Three, these idealizations, strictly speaking, cannot be used to get the approximate solutions to the dynamically holistic systems (i.e., the real systems in the world). Dynamically holistic systems are singular perturbation problems which do not satisfy the above commutation condition: That is, for these systems, the solutions to the unperturbed (or approximate) equations are not the approximate solutions of them. This means that we cannot directly export the solutions of the unperturbed equations to the real systems. These approximate solutions, obtained by singular perturbation methods, are very different from the solutions of the unperturbed systems. For example, the approximate solution (to the first order) to the Duffing equation is $x(t) = a\cos(t + \frac{3}{8}\varepsilon t a^2 + \beta_0)$, whereas the solution for the unperturbed system is $x = a\sin t$. This is because, in the case of multiple scales method, after we introduce multiple scales variables in the equations of the real system, the first term in the solution (or the first order approximate solution) is no longer the solution of the unperturbed system but a solution written as $x_0 = a(T,T_1...)\cos[t+\beta(T,T_1...)]$, which (especially the functions $a(T,T_1...), \beta(T,T_1...)$ needs equations at higher orders (e.g., equations at the order of $\varepsilon, \varepsilon^2 etc.$) to get determined. These equations at higher orders involve perturbation terms. For example, the equation at the order of ε is $\frac{\partial^2 x_1}{\partial t^2} + x_1 = -2\frac{\partial^2 x_0}{\partial t \partial T} - x_0{}^3$ where the term $x_0{}^3$ is the perturbation term.

---

[29] This point is suggested by my supervisor Rueger, who also described it in Rueger, 2001.

As I have shown in the examples in the last section, the approximate solutions contain the effects of the average parts of the perturbations. The approximate solution (to a real system) is the (exact) solution to the equations of motion of the averaged system, i.e. the equations of motion for the unperturbed system plus averaged perturbation terms (in the averaging method). For example, the general equation of motion for a perturbed Hamiltonian system in terms of action-angle variables is:

$$\dot{I} = -\varepsilon \frac{\partial H_1}{\partial \varphi} \quad , \quad \dot{\varphi} = \frac{\partial H_0}{\partial I} + \varepsilon \frac{\partial H_1}{\partial I}$$

where $H = H_0(I) + \varepsilon H_1(I,\varphi,\varepsilon)$, H_0 is the unperturbed Hamiltonian and H_1 is the perturbing Hamiltonian. The averaged Hamiltonian for this system can be written as $\overline{H} = H_o(J) + \varepsilon \overline{H}_1(J)$ where $\overline{H}_1(J)$ is the average part of H_1.[30] So the oscillatory part of the perturbation H_1 is removed from the averaged system. The equation for the averaged system is: $\dot{J} = 0$, $\dot{\psi} = \frac{\partial H_0}{\partial J} + \varepsilon \frac{\partial \overline{H}_1}{\partial J}$. In contrast, the equation for the unperturbed system is: $\dot{I} = 0$, $\dot{\varphi} = \frac{\partial H_0}{\partial I}$.

---

[30] Here the variable J (as well as the variable ψ below) represent the variables of the averaged system which has the averaged Hamiltonian, replacing by I and φ respectively.

So the approximate solution to the real system can also be seen as the exact solution to the integrable averaged system with the averaged Hamiltonian which is the sum of the Hamiltonian of the unperturbed system plus the Hamiltonian of the average part of the perturbation. This implies that singular perturbation methods are not like regular perturbation methods which divide a real system into an unperturbed system and perturbations, rather, they divide a real system into an integrable (averaged) system and a nonintegrable system. For regular perturbation methods, idealizations mean to remove (or reduce) the perturbation terms in the equations of motion of the real systems, which is how the philosophers who take regular perturbation methods as the basis for idealizations and exportations think of idealization. But singular perturbation methods show that in order to get idealized models that are exportable (as approximations) to the real systems, we have to idealize real systems by removing only the oscillatory parts of the perturbations in the Hamiltonians (or the equations of motion), consequently removing only the effects of the oscillatory parts of perturbations in the solutions of the real systems. So instead of taking the unperturbed system as the idealized model, singular perturbation methods suggest that we take the averaged system as the idealized model.

It seems that if we idealize the equation of motion for a real system to that of the averaged system, we can get an approximate solution from the approximate equation. In this way, we seem to recover the commutation of "idealizing" the equation and "idealizing" the solution. However, we should note that the recovery succeeds only when the singular perturbation methods work (i.e., we can get a uniform asymptotic solution to the real system from the averaged equation). Also, the recovery requires that we idealize the equation in a certain way (i.e., eliminate only the oscillatory part of perturbations) in order to get the correct idealized solution (as a valid approximation). So the commutation is severely restricted.

So the singular perturbation methods imply that we need a new notion of idealized models and thus a different view of the process of idealization for dynamically holistic systems. For these systems, idealized models have to be the averaged systems that include or have the averaged interactions or perturbations. Therefore, idealization in dynamically holistic cases does not take away all the perturbations but omits only the oscillatory part of the perturbations (in the equation of a real system), or truncates only the small correction part of the solution that results from the oscillatory part of the perturbations. To distinguish these new idealized models (as averaged systems) from the traditional idealized models (as unperturbed systems), I will call these new ones "modified" idealized models. Only these modified idealized models can be exported as approximations to the behavior of the real systems. Since the scientist in the epistemic situation of idealization has the real descriptions, he can see that idealized models need to take averaged interactions into account; singular perturbation methods provide the tool for arriving at properly idealized models. So in this case, singular perturbation methods are a recipe for the scientists for how to idealize real systems.

Can singular perturbation methods also provide justification for scientists' exporting of the modified idealized models to the real systems? Suppose we know the description of the real system and try to justify the claim that the description of the modified idealized model can be exported as an approximate description of the real system, -- can singular perturbation methods then justify the exportation?

In the last section I have shown that the asymptotic solutions obtained from singular perturbation methods are valid approximations, that is, they are *quantitatively* valid approximations. As Arnold puts it, the notion of asymptoticness means that after truncating the series at the terms of order ε^n, the remainder is approximately true with the error of the order ε^{n+1} (for initial conditions for which no resonances occur in the solution) (Arnold 1983, p.150). The asymptoticness of the solution guarantees that any truncation of the expansion of the solution of the real system is a quantitative approximation; even the first term alone is quantitatively approximate with the error of the order of ε. This means that the modified idealized models can be exported to real systems as quantitatively approximate descriptions.

Some may think the modified idealized models may not be *qualitatively* similar to the real systems and are thus not justifiably exportable. As Liu (1999) argues, idealized models must qualitatively approximate the real systems. Merely quantitatively close descriptions cannot be used as idealized models. Rueger and Sharp also point out that good idealized models must be qualitatively similar to the true description (Rueger and Sharp, 1998, p.10). This means that for the modified idealized models to be exportable we need to show that these models are qualitatively similar to the true solutions of dynamically holistic systems.

From the example of the van der Pol problem, we can see that the modified idealized model is structurally similar to that of the true description of the behavior of the perturbed system within a wide domain (of the variable) where the asymptotic solution is valid. (See Figures 4 and 5.) Although the phase space portrait of the unperturbed idealized system is very different from that of the perturbed system, the introduction of multiple scales superimposes a slow drift process on the fast oscillation process of the idealized unperturbed system and makes the structure of the uniform asymptotic solution similar to that of the perturbed system. When the solution of an unperturbed system uniformly approximates that of a regularly perturbed system, the two systems have the same topological structure; similarly, when the modified idealized model uniformly approximates the behavior of the real system (under the conditions specified in Section 4.2), it also has the same topological structure as the latter.

But we should note that the structural similarity between the modified idealized model and the real system does not mean that the unperturbed model is structurally stable against perturbations. The similarity, rather, is between the trajectories of the (averaged) perturbed system and those of the real system that the former approximate. So the modified idealized models are not only quantitatively close but also qualitatively similar to the behavior of the dynamically holistic systems.

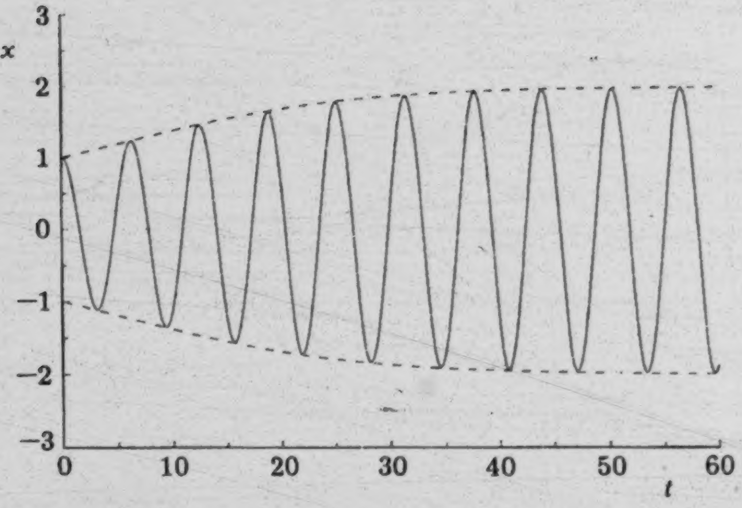

Fig. 7.1 The continuous curve is a numerical solution $x(t;\epsilon)$ of the initial value problem for a van der Pol oscillator with $\epsilon = 0.1$. The dashed curves give the asymptotic predictions for the amplitude, i.e. $\pm R(T)$. The agreement is good at even larger values of the small parameter, as is typical of this class of asymptotic analysis.

Figure 4

From E. J. Hinch, 1981, P.109, figure 7.1

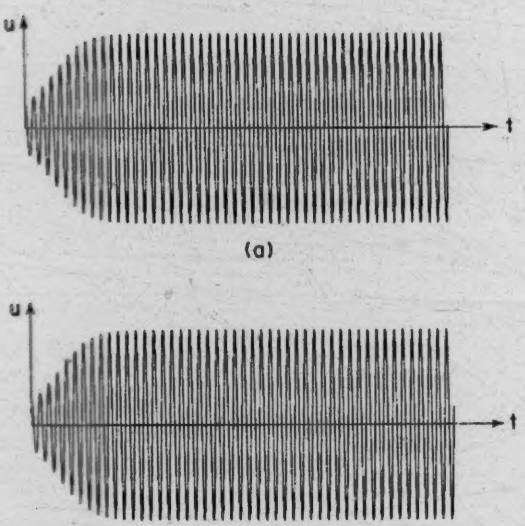

Figure 6-3. Comparison of the approximate solution (*b*) with the exact solution (*a*) for $u(0) = 0.5$, $\dot{u}(0) = 0$, and $\epsilon = 0.1$.

Figure 5

From A. H. Nayfeh, 1981, P.155, figure 6-3

Therefore, the modified idealized models are justified as exportable to the dynamically holistic systems. Still, some may not agree because the behavior of the modified idealized models only asymptotically approximates the behavior of the real systems but may not converge to the latter. This means that we may not be able to reach the true description of the behavior of the real system by adding back the oscillatory part of the perturbation.

Surely it would be best if we could get asymptotic and convergent de-idealization series, and theoretically we could get a KAM expansion which is both asymptotic and convergent. However, as Lichtenberg and Lieberman note, the KAM method "has not been widely applied to practical perturbation calculations. …the method is generally too cumbersome to use in this way." (Lichtenberg and Liberman, 1983, p.140). So practically scientists use multiple scales or averaging methods to get asymptotic approximations for real systems.

Since the asymptotic approximations are valid only under certain conditions, we have to use the exportation of the modified idealized models with caution: we have to restrict the exportation to the generic regular motion (i.e., most initial conditions where no resonances are possible in the solutions to the order of ε^{n+1} if the asymptotic solutions are the order of ε^n) of the generic systems (whose unperturbed systems are non-degenerate and the perturbations of these systems are small enough); and we can only take the first n idealizations (the nth idealization means the truncation of the asymptotic expansion after the first n terms) as approximate (if the solutions are uniformly asymptotic to the order of ε^n); and if we introduce only n-1 new scale variables, e.g., $T_1 = \varepsilon^2 t, \ldots T_{n-1} = \varepsilon^n t$, we can take the asymptotic description as approximate only in the domain: $0 \le t \le \varepsilon^{-n}$. Since the uniform asymptotic expansions are guaranteed only to be asymptotic (not convergent), they may diverge after several terms, which means that these asymptotic solutions obtained by using the multiple scales or averaging methods may never reach the exact solutions (for any initial conditions) of the dynamically holistic systems.[31] However, as long as the asymptoticness does not break down, valid asymptotic solutions are good approximations.[32]

McMullin might not agree with this new account of idealization. He might argue that if the asymptotic series is divergent, it means that we did not get the (causal) structure or pattern of the real systems right, and therefore even if the asymptotic solution can be quantitatively very close to the true solution, it cannot be taken as even approximately true. For McMullin, to get the approximate truth is not just to get the quantity close but to get the causal structure right. Only when the true descriptions converge to the idealizations, do the idealizations idealize the real structure of the systems and therefore are approximations. The divergent asymptotic solutions for McMullin do not approximate the true structure (McMullin, 1985, pp.261-262) because they suggest a structure for which the initial adding back of the idealized factors improves the predictions, but the subsequent adding back of factors makes the predictions worse again. Moreover, the divergence of the asymptotic solutions means that they do not converge to anything or they represent a constructed object that actually does not exist (Arnold, 1983, p.178). So the divergent asymptotic solutions cannot be approximations to the true structure of the real objects.

---

[31] However, theoretically we could get convergent and asymptotic expansions with multiple scales from the KAM method, which means that theoretically we could reach the true solutions of the nondegenerate systems with small perturbations for their non-resonant initial conditions through the KAM de-idealization series.

[32] Although the KAM theorem provides support for the validity of the asymptotic solutions since it shows the asymptotic solution is generally right in describing the generic behavior of the real systems (i.e., for most initial conditions of the generic nondegenerate systems), uniform asymptotic solutions are themselves justified approximate solutions.

To this objection, my answer is that although the uniform asymptotic expansions could be divergent, it does not follow that they cannot describe the true structure of the real systems. As discussed in Chapter One, the structure of the dynamics of a system is represented by the topological structure of its phase portrait. As I have shown above, within some range of the variables, and for the generic initial conditions, the structure of the modified idealized system is similar to that of the real system. In this sense the modified idealized model does capture the true structure of the real system. The notion of true structure of real systems for McMullin seems to mean the causal structure.[33] Since causal factors of a system are described in the equations of motion, for McMullin getting the structure right in idealizations means to get the (idealized) equations right, that is, the equations of real systems would converge to the idealized equations when the perturbations approach zero. However, as we have discussed, although the *equations* of singularly perturbed real systems converge to the equations of idealized unperturbed systems, the *solutions* for the former do not converge to the solutions for the latter. But because McMullin implicitly assumes that the operations of idealizing equations and of idealizing commute, his requirement of "getting the structure right" for idealized models also implies that the solutions for the real systems converge to the solutions for the idealized models as perturbations approach zero. Although singular perturbation methods "restore" the commutation in the sense that the solution of the approximate equation, i.e., the equation of the modified idealized model, approximates the solution of the real system, the latter does not converge to the former.

---

[33] McMullin defines the term "structure" as "a set of constituent entities or processes and the relationship between them" and he calls the structure "the cause of the feature being explained". (McMullin, 1978, p.139.)

In summary, the accounts of McMullin, Cartwright, and Rueger and Sharp justify only the exportation of the traditional idealized models, i.e., the unperturbed systems to regularly perturbed systems. They cannot justify the exportation of idealized models to real systems which are dynamically holistic. Singular perturbation methods show that we have to idealize real systems differently (from simply eliminating all perturbations) to get modified idealized models that are exportable to them. The justification of the exportation of the modified idealized models to the real systems is based on the fact that the former approximate the behavior of the latter.

The above new account of exportation based on the notion of modified idealized model may be regarded as a modified version of McMullin's theory: the modified idealized models can be exported as approximations to the real systems because the idealized oscillatory perturbations are small corrections. The behavior of the real system can be seen as the behavior of the idealized model plus small corrections. As I have shown, the modified idealized models also have the same structure within certain domains as the real systems. Thus, we may say local structural similarity can justify the exportation of the qualitative behavior of the modified idealized models to the real systems. This justification of exportation may be seen as a modified theory of Rueger and Sharp. If we take the modified idealized models as describing a part of the behavior of the real systems since the motion (or solutions) of the real systems are the motion of the modified idealized systems plus the correction terms, we may say they are justified to be exportable as a part of the real systems. This justification may be seen as a revised theory of Cartwright although the modified idealized models do not describe a parameter stable part of the behavior of the real systems.

(c) A new account of traditional idealized models

What I have shown above is that modified idealized models can justifiably be exported to real systems. What about the traditional idealized models? Can they be exported to real systems in *some* way? If not, what do they describe?

Singular perturbation methods show that traditional idealized models cannot be approximations to the real systems, which invalidates McMullin's view that idealized models are approximate descriptions of the real systems. Neither are the idealized models in general structurally stable, which renders incorrect Rueger and Sharp's taking idealized models as describing the structure or qualitative behavior of the real systems. Do the idealized laws perhaps describe some kind of stable tendencies in Cartwright's sense in the dynamically holistic world?

Obviously, traditional idealized models cannot be exported as parameter stable tendencies as real systems are not regular perturbation problems. Can a traditional model nevertheless be exported as a part, in some sense, of a real system?

The answer for this question would also provide an account of what the traditional idealized models describe.

To answer the question we need to see whether the solution of the unperturbed system constitutes a part of the solution of the real system. As I have shown in the examples in Section 1 of this chapter, the solution of the unperturbed system (i.e., the traditional idealized models) is always a one scale part of the approximate solution of the dynamically holistic system. Moreover, this relation is not accidental but general. This is because the solutions of the unperturbed systems should be some kind of limits of the asymptotic approximate solutions. That is, for the asymptotic approximations to be legitimate solutions of the real systems, they are not only required to approximate the solutions of the real systems, but also required to become the solutions of the unperturbed systems when the effects of perturbations are reduced to zero. In other words, since the solutions of the unperturbed systems can be seen as the solutions of the real systems when there are no perturbations in them, the approximate solutions should go over to the solutions of the unperturbed systems when the effects of perturbations (which are the contributions in terms of the slow time scales) in the approximate solutions go to zero. This means that the solutions of the unperturbed systems or the traditional idealized models are the limits of the uniform asymptotic solutions when $\varepsilon \to 0$; $\varepsilon t \to 0$; $\varepsilon^2 t \to 0 \cdots$. For example, the uniform asymptotic solution of the Duffing equation will become the solution of the unperturbed system in the limit as $\varepsilon \to 0$; $\varepsilon t \to 0$:

$$\lim_{\varepsilon \to o, \varepsilon t \to 0,} a\cos(t + \frac{3}{8}\varepsilon t a^2 + \beta_0) + O(\varepsilon) = a\cos(t + \beta_0))$$

It might seem that since the idealized solution of the unperturbed system is always a one scale part of the solution of the real system, these one scale parts might be understood as describing one scale tendencies of real systems, in a sense akin to Cartwright's. However, the affinity with Cartwright's notion of tendency is imperfect:

(1) The solution of the idealized unperturbed system is not a separable one scale part of the uniform asymptotic approximation of the real system. As we can see from the uniform asymptotic solution of the real system (e.g., the solution of the Duffing equation: $x = a\cos(t + \frac{3}{8}\varepsilon t a^2 + \beta_0) + O(\varepsilon)$) the solution of the idealized unperturbed system ($x = a\cos(t + \beta_0)$) does not constitute a separable functional part in the solution of the real system. This is because both its amplitude and phase can be modified and the contributions from different scales are mingled or entwined together so that it is not possible to separate them as independent functions (e.g., the solution of the real system cannot be written as $f(t)g(\varepsilon t)h(\varepsilon^2 t)\cdots$). The solution of the idealized unperturbed system is a one scale part of the solution of the real system only when the contributions from other scales are eliminated. This means that we cannot directly export the solution of the idealized unperturbed system to the solution of the real system as a separable functional part or a separable term of it (whereas Cartwright supposes her stable tendency can be exported in such ways). So the one scale tendency is not exportable by itself. Without the uniform asymptotic approximation, even if we could export the solution of the idealized

unperturbed system as describing the one scale motion of the real system, it does not help us to know the behavior of the real system.

(2) Although singular perturbation methods show that the solution or motion of the unperturbed system always constitutes a (one scale) part of the solution of the real nonseparable system, it does not mean that it is a stable tendency in Cartwright's sense. This is because the idealized solution of the unperturbed system does not describe the stable effect (in terms of stable parameter or functional form) of an atomistic factor. The solution of the ideal system cannot be taken as describing the effect of an atomistic or separable factor because if we describe all the factors and their effects of the dynamically holistic systems all at the same scale, the effects (at the level of solution) of these factors are not separable. Even if we describe these factors (say Hamiltonians) and their effects (in the solution) in terms of multiple scales, since the scales are entangled or entwined in the solution, we still cannot take the idealized solution (i.e., the solution of the unperturbed system) as describing the separable effect of an atomistic factor.

Conclusion

Singular perturbation methods show that the problem of justifying the exportation of idealized models to dynamically holistic systems to get approximate descriptions of them can be solved by a new way of idealizing real systems. Instead of taking the solutions of the unperturbed systems as idealized models, to obtain properly idealized models, i.e., models that approximate the behavior of the real systems, we have to take the averaged perturbation into account to form modified idealized models. So it is the averaged systems (as modified idealized models) rather than the unperturbed systems that approximate the behavior of the real systems. The unperturbed systems which are traditionally taken as idealized models that are supposed to be directly exportable to the real systems, turn out to be exportable only indirectly, that is, only exportable as one scale parts of the modified idealized models. Although we may take these traditional idealized models as describing one scale tendencies, they are essentially different from Cartwright's notion of stable tendencies.

Our theory of exportation based on singular perturbation theory indicates that we can export (modified) idealized models to dynamically holistic systems without their being Nickles reducible to these models or the traditional idealized models. Although the Nickles nonreducibility of these holistic systems to the modified idealized models restricts the exportation to a certain domain of the variable, the new exportation theory does not have the traditional *ceteris paribus* problem because the introduction of multiple scales in these modified idealized models enables them to be exportable to as wide a domain of the variable as we choose.

With the introduction of the notion of modified idealized models, we can say that although the dynamically holistic world described by classical mechanics cannot be Nickles reduced to the traditional idealized models, not even Nickles reduced to the modified idealized models, its behavior can be approximated by the superimposition of the behaviors at different scales. The modified idealized models provide us with a way of "carving" the dynamically holistic world, that is, we can carve the dynamically holistic world into behaviors at different scales. In this sense, we may say the dynamically holistic world is approximately "dynamically atomistic" with respect to the motions at different scales or levels.

Appendix

The invariance of the Hamiltonian under canonical transformation is due to the covariance of generalized momenta and the contravariance of generalized coordinates. Generalized momenta can be expressed as gradients of a scalar field:

$$p_i = \frac{\partial W}{\partial q_i}$$

where *W* is the action integral.

This shows that the generalized momenta transform as covariant variables (whereas generalized velocities, which are tangent vectors and transform as contravariant variables, cannot be expressed as the gradient of a scalar field). Because of the covariance of the generalized momenta and the contravariance of the dq_i under the coordinate change from $\boldsymbol{q} = q\ (q_1, q_2, \ldots q_n\,)$ to $\boldsymbol{Q} = \boldsymbol{Q}\ (Q_1, Q_2, \ldots Q_n\,)$ (see Tabor, 1989, pp.81-84),

$$\int p_1 dq_1 + p_2 dq_2 + \ldots + p_n dq_n$$

is an integral invariant, which means that it is invariant under the canonical transformation:

$$P_i = P_i(q_1,\cdots q_n, p_1,\cdots p_n)$$
$$Q_i = Q_i(q_1,\cdots q_n, p_1,\cdots p_n)$$

That is,

$$\int p_1 dq_1 + p_2 dq_2 + ... + p_n dq_n = \int P_1 dQ_1 + P_2 dQ_2 + ... + P_n dQ_n$$

This integral invariance implies the invariance of the Hamiltonian form of the differential equations, that is, the differential equations in terms of the new variables P_i, Q_i still have the Hamiltonian form:

$$\dot{Q}_i = \frac{\partial H'}{\partial P_i}$$
$$\dot{P}_i = -\frac{\partial H'}{\partial Q_i} \qquad i = 1,2,\cdots n$$

where $H' = H'(Q_1,Q_2,\cdots Q_n, P_1, P_2,\cdots P_n)$

is the transformed Hamiltonian (see Whittaker, 1952, pp.305-306). So H and H' are the same function expressed in different sets of canonically conjugate variables (i.e., generalized coordinates and generalized momenta) (see ter Haar, 1971, p.100). So the Hamiltonian and Hamilton's equation of motion are invariant under canonical transformation.

Bibliography

Albert, D. 1992, *Quantum Mechanics and Experience.* Harvard University Press.

Arnold, V. I. 1978, *Mathematical Methods of Classical Mechanics.* Spinger-Verlag.

Arnold, V.I. (ed.), 1983, *Dynamical Systems III.* Springer-Verlag.

Batterman, R.W. 1995, "Theories between Theories", *Synthese* 103, pp.171-201.

Batterman, R. W. 2003, "Intertheory Relations in Physics." *Stanford Encyclopedia of Philosophy, http://plato.stanford.edu/entries/physics-interrelate*

Birkhoff, G. 1927, *Dynamical Systems.* American Mathematical Society.

Cartwright, N. 1989, *Nature's Capacities and their Measurement.* Claredon Press.

Cartwright, N. 1983, *How the Laws of Physics Lie.* Clarendon Press.

Cartwright, N. 1991, "Can Wholism Reconcile the Inaccuracy of Theory with the Accuracy of Prediction?" *Synthese* 89, pp. 3-13.

Cartwright, N. 1999, *The Dappled World: A study of the Boundaries of Science.* Cambridge University Press.

Cohen, E.G. D. (ed.), 1980, *Fundamental Problems in Statistical Mechanics V.* North-Holland Publishing Company.

Cole, J. D. 1968, *Perturbation Methods in Applied Mathematics.* Blaisdell Publishing Company.

Corben, H.C. and Stehle, P. 1960, *Classical Mechanics.* Second edition. Robert E. Krieger Publishing Company.

Flaschka, H., Newell, A.C. and Tabor, M. 1991, "Integrability", in V. E. Zakharov (ed.) *What is Integrability?* Springer-Verlag, pp. 348-372.

Giacaglia, E. E. O. 1972, *Perturbation Methods in Non-Linear Systems.* Springer-Verlag.

Glendining, P. 1994, *Stability, Instability and Chaos.* Cambridge University Press.

Goldstein, H. 1950, *Classical Mechanics.* Addison-Wesley Publishing Company.

Groesberg, S.W. 1968, *Advanced Mechanics.* John Wiley and Sons.

Gukenheimer, J. and Holmes, P. 1983, *Nonlinear Oscillations, Dynamical Systems, and Bifurcations of Vector Fields.* Springer-Verlag.

Healey, R. 1991, "Holism and Nonseparability", *Journal of Philosophy* 88, pp.393-421.

Helleman, R.H.G. 1980, "Self-Generated Chaotic Behavior in Nonlinear Mechanics", in E.G. D. Cohen (ed.) 1980, pp.165-233.

Hinch, E. J. 1991, *Perturbation Methods.* Cambridge University Press.

Holmes, M. 1999, "The Method of Multiple Scales" in J. Cronin and R. O'Malley (eds.), *Proceedings of Symposia in Applied Mathematics.* Vol.56.

Kim, J. 1998, "Reduction, problems of", in E. Craig (ed.), *Routledge Encyclopedia of Philosophy,* http://www.rep.routledge.com/article/Q089SECT2

Kevorkian, J. and J. D. Cole 1996, *Multiple Scale and Perturbation Methods.* Springer-Verlag.

Kronz, F. 1998, "Nonseparability and Quantum Chaos", *Philosophy of Science* 65, pp. 50-75.

Kronz, F. and J. Tiehen, 2002, "Emergence and Quantum Mechanics", *Philosophy of Science* 69, pp. 324-347.

Lichtenberg, A.J. and M. A. Lieberman 1983, *Regular and Stochastic Motion.* Springer Verlag.

Liu, C. 1999, "Approximation, Idealization and Laws of Nature", *Synthese* 118, pp. 229-256.

Maudlin, T. 1998, "Part and Whole in Quantum Mechanics", in E. Castellani (ed.), *Interpreting Bodies.* Princeton University Press, pp. 46-60.

McMullin, E. 1985, "Galilean Idealization", *Studies in History and Philosophy of Science* 16, pp. 247-273.

McMullin, E. 1978, "Structural Explanation", *American Philosophical Quarterly* 15, pp.139-147.

Meyer, K. R. and G. R. Hall 1991, *Introduction to Hamiltonian Dynamical Systems and the N-Body Problem.* Springer-Verlag.

Morrison, M. 1995, "Capacities, Tendencies and the Problem of Singular Causes", *Philosophy and Phenomenological Research* 55, pp. 163-168.

Nagel, E. 1961, *The Structure of Science*. Routledge and Kegan Paul.

Nayfeh, A. H. 1981, *Introduction to Perturbation Techniques.* John Wiley and Sons.

Nickles, T. 1973, "Two Concepts of Intertheoretic Reduction", *Journal of Philosophy* 70, pp.181-201.

O'Malley, R.E. Jr. 1991, *Singular Perturbation Methods for Ordinary Differential Equations.* Springer-Verlag.

Rohrlich, F. 1988, "Pluralistic Ontology and Theory Reduction in the Physical Sciences", *British Journal for the Philosophy of Science* 39, pp. 295-312.

Rueger, A. 2000, "Physical Emergence, Diachronic and Synchronic", *Synthese* 124, pp. 297-322.

Rueger, A. 2001, "Explanation at Multiple Scales", *Minds and Machines* 11, pp. 503-520.

Rueger, A. 2004, "Reduction, Autonomy, and Causal Exclusion Among Physical Properties", *Synthese* 139, pp. 1-21.

Rueger, A. and D. Sharp 1998, "Metaphysical Presuppositions of Scientific Practice: 'Atomism' vs. 'Wholism', *Canadian Journal of Philosophy* 28, pp. 1-20.

Saletan, E. and A. Cromer 1971, *Theoretical Mechanics.* John Wiley and Sons.

Scheibe, E. 1995, "Laws and Theories: Generality vs. Coherence" in F. Weinert (ed.), *Laws of Nature.* DeGruyter. pp.208-226.

Shirer, H. 1987, *Nonlinear Hydrodynamic Modeling.* Springer-Verlag.

Sklar, L. 1967, "Types of Inter-theoretic Reduction", *British Journal for the Philosophy of Science* 18, pp.109-124.

Smith, D.R. 1985, *Singular Perturbation Theory.* Cambridge University Press.

Sposito, G. 1976, *An Introduction to Classical Dynamics.* John Wiley and Sons.

Tabor, M. 1989, *Chaos and Integrability in Nonlinear Dynamics: an Introduction.* John Wiley and Sons.

ter Haar, D. 1971, *Elements of Hamiltonian Mechanics.* Pergamon Press.

Whittaker, E.T. 1952, *A Treatise on the Analytical Dynamics of Particles and Rigid Bodies.* Cambridge University Press.

Zeeman, E. C. 1988, "Stability of Dynamical Systems", *Nonlinearity* 1, pp.115-155.